

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



BI

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification⁵ : C12N 15/12, 15/63, 15/64 C12N 5/10, C07K 13/00 A01N 33/00, A61K 37/00</p>	<p>A1</p>	<p>(11) International Publication Number: WO 92/17585 (43) International Publication Date: 15 October 1992 (15.10.92)</p>
<p>(21) International Application Number: PCT/US92/02741 (22) International Filing Date: 6 April 1992 (06.04.92) (30) Priority data: 681,880 5 April 1991 (05.04.91) US (60) Parent Application or Grant (63) Related by Continuation US 681,880 (CIP) Filed on 5 April 1991 (05.04.91) (71) Applicant (for all designated States except US): THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK [US/US]; Broadway and West 116th Street, New York, NY 10027 (US).</p>		<p>(72) Inventors; and (75) Inventors/Applicants (for US only): BUCK, Linda, B. [US/US]; 100 Haven Avenue, New York, NY 10032 (US). AXEL, Richard [US/US]; 445 Riverside Drive, New York, NY 10027 (US). (74) Agent: WHITE, John, P.; Cooper & Dunham, 30 Rockefeller Plaza, New York, NY 10112 (US). (81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), MC (European patent), NL (European patent), SE (European patent), US. Published With international search report.</p>
<p>(54) Title: ODORANT RECEPTORS AND USES THEREOF</p> <p>(57) Abstract</p> <p>The invention provides an isolated nucleic acid, e.g. cDNA encoding an odorant receptor. The invention further provides expression vectors containing such nucleic acid. Also provided is a purified protein encoding an odorant receptor, with the aforementioned expression vectors and the resulting transformed cell. The invention also provides methods of identifying odorant ligands and of identifying odorant receptors. The invention further provides methods of developing fragrances, of identifying appetite suppressant compounds, of controlling appetite, of controlling pest populations, of promoting and inhibiting fertility, and of detecting odors.</p>		



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	RU	Russian Federation
CG	Congo	KP	Democratic People's Republic of Korea	SD	Sudan
CH	Switzerland	KR	Republic of Korea	SE	Sweden
CI	Côte d'Ivoire	LI	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
DE	Germany	MC	Monaco	TC	Togo
DK	Denmark			US	United States of America

ODORANT RECEPTORS AND USES THEREOFBackground of the Invention

5 This application is a continuation-in-part of U.S. Serial No.681,880, filed April 5, 1991, the contents of which are hereby incorporated by reference.

10 Throughout this application, various publications are referenced by Arabic numerals within parentheses. Full citations for these publications may be found at the end of the specification immediately preceding the claims. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in
15 order to more fully describe the state of the art as known to those skilled therein as of the date of the invention described and claimed herein.

20 In vertebrate sensory systems, peripheral neurons respond to environmental stimuli and transmit these signals to higher sensory centers in the brain where they are processed to allow the discrimination of complex sensory information. The delineation of the peripheral mechanisms by which environmental stimuli are transduced into neural information
25 can provide insight into the logic underlying sensory processing. Our understanding of color vision, for example, emerged only after the observation that the discrimination of hue results from the blending of information from only three classes of photoreceptors (1, 2, 3, 4). The basic
30 logic underlying olfactory sensory perception, however, has remained elusive. Mammals possess an olfactory system of enormous discriminatory power (5, 6). Humans, for example, are thought to be capable of distinguishing among thousands of distinct odors. The specificity of odor recognition is
35 emphasized by the observation that subtle alterations in the molecular structure of an odorant can lead to profound

-2-

changes in perceived odor.

The detection of chemically distinct odorant presumably results from the association of odorous ligands with specific receptors on olfactory neurons which reside in a specialized epithelium in the nose. Since these receptors have not been identified, it has been difficult to determine how odor discrimination might be achieved. It is possible that olfaction, by analogy with color vision, involves only a few odor receptors, each capable of interaction with multiple odorant molecules. Alternatively, the sense of smell may involve a large number of distinct receptors each capable of associating with one or a small number of odorant. In either case, the brain must distinguish which receptors or which neurons have been activated to allow the discrimination between different odorant stimuli. Insight into the mechanisms underlying olfactory perception is likely to depend upon the isolation of the odorant receptors, and the characterization of their diversity, specificity, and patterns of expression.

The primary events in odor detection occur in a specialized olfactory neuroepithelium located in the posterior recesses of the nasal cavity. Three cell types dominate this epithelium (Figure 1A): the olfactory sensory neuron, the sustentacular or supporting cell, and the basal cell which is a stem cell that generates olfactory neurons throughout life (7, 8). The olfactory sensory neuron is bipolar: a dendritic process extends to the mucosal surface where it gives rise to a number of specialized cilia which provide an extensive, receptive surface for the interaction of odors with olfactory sensory neurons. The olfactory neuron also gives rise to an axon which projects to the olfactory bulb of the brain, the first relay in the olfactory system. The axons of the olfactory bulb neurons, in turn, project to

-3-

subcortical and cortical regions where higher level processing of olfactory information allows the discrimination of odors by the brain.

5 The initial events in odor discrimination are thought to involve the association of odors with specific receptors on the cilia of olfactory neurons. Selective removal of the cilia results in the loss of olfactory response (9). Moreover, in fish, whose olfactory system senses amino acids
10 as odors, the specific binding of amino acids to isolated cilia has been demonstrated (10, 11). The cilia are also the site of olfactory signal transduction. Exposure of isolated cilia from rat olfactory epithelium to numerous odorant leads to the rapid stimulation of adenylyl cyclase
15 and elevations in cyclic AMP (an elevation in IP3 in response to one odorant has also been observed) (12, 13, 14, 15). The activation of adenylyl cyclase is dependent on the presence of GTP and is therefore likely to be mediated by receptor-coupled GTP binding proteins (G-proteins) (16).
20 Elevations in cyclic AMP, in turn, are thought to elicit depolarization of olfactory neurons by direct activation of a cyclic nucleotide-gated, cation permeable channel (17, 18). This channel is opened upon binding of cyclic nucleotides to its cytoplasmic domain, and can therefore
25 transduce changes in intracellular levels of cyclic AMP into alterations in the membrane potential.

These observations suggest a pathway for olfactory signal transduction (Figure 1B) in which the binding of odors to
30 specific surface receptors activates specific G-proteins. The G-proteins then initiate a cascade of intracellular signalling events leading to the generation of an action potential which is propagated along the olfactory sensory axon to the brain. A number of neurotransmitter and hormone
35 receptors which transduce intracellular signals by

-4-

activation of specific G-proteins have been identified. Gene cloning has demonstrated that each of these receptors is a member of a large superfamily of surface receptors which traverse the membrane seven times (19, 20). The pathway of olfactory signal transduction (Figure 1B) predicts that the odorant receptors might also be members of this superfamily of receptor proteins. The detection of odors in the periphery is therefore likely to involve signalling mechanisms shared by other hormone or neurotransmitter systems, but the vast discriminatory power of the olfactory system will require higher order neural processing to permit the perception of individual odors. This invention address the problem of olfactory perception at a molecular level. Eighteen different members of an extremely large multigene family have been cloned and characterized which encodes seven transmembrane domain proteins whose expression is restricted to the olfactory epithelium. The members of this novel gene family encode the individual odorant receptors.

20

-5-

SUMMARY OF THE INVENTION

The invention provides an isolated nucleic acid, e.g. a DNA and cDNA molecule, encoding an odorant receptor. The invention further provides expression vectors containing such nucleic acid. Also provided by the invention is a purified protein encoding an odorant receptor. The invention further provides a method of transforming cells which comprises transfecting a suitable host cell with a suitable expression vector containing the nucleic acid encoding the odorant receptor.

The invention also provides methods of identifying odorant ligands and of identifying odorant receptors. The invention further provides methods of developing fragrances, of identifying appetite suppressant compounds, of controlling appetite. The invention also provides methods of controlling insect and other animal populations. The invention additionally provides a method of detecting odors such as the vapors emanating from Cocaine, Marijuana, Heroin, Hashish, Angel Dust, gasoline, decayed human flesh, alcohol, gun powder explosives, plastic explosives, firearms, poisonous or harmful smoke, or natural gas.

-6-

Description of the Figures

Figure 1. The Olfactory Neuroepithelium and a Pathway for Olfactory Signal Transduction. A. The Olfactory Neuroepithelium. The initial event in odor perception occurs in the nasal cavity in a specialized neuroepithelium which is diagrammed here. Odors are believed to interact with specific receptors on the cilia of olfactory sensory neurons. The signal generated by these initial binding events are propagated by olfactory neuron axons to the olfactory bulb. B. A Pathway of Olfactory Signal Transduction. In this scheme, the binding of an odorant molecule to an odor-specific transmembrane receptor leads to the interaction of the receptor with a GTP-binding protein ($G_{S[olf]}$). This interaction, in turn, leads to the release of the GTP-coupled α -subunit of the G-protein, which then stimulates adenylyl cyclase to produce elevated levels of cAMP. The increase in cAMP opens nucleotide-gated cation channels, thus causing an alteration in membrane potential.

20

Figure 2. A PCR Amplification Product Containing Multiple Species of DNA. cDNA prepared from olfactory epithelium RNA was subjected to PCR amplification with a series of different primer oligonucleotides and the DNA products of appropriate size were isolated, further amplified by PCR, and size fractionated on agarose gels (A) (For details, see text). Each of these semipurified PCR products was digested with the restriction enzyme, Hinf I, and analyzed by agarose gel electrophoresis. Lanes marked "M" contain size markers of 23.1, 9.4, 5.6, 4.4, 2.3, 2.0, 1.35, 1.08, 0.87, 0.60, 0.31, 0.28, 0.23, 0.19, 0.12 and 0.07kb. (B). Twenty-two of the 64 PCR products that were isolated and digested with Hinf I are shown here. Digestion of one of these, PCR 13, yielded a large number of fragments whose sizes summed to a value much greater than that of the undigested PCR 13

35

-7-

DNA, indicating that PCR 13 might contain multiple species of DNA which are representatives of a multigene family.

Figure 3. Northern Blot Analysis with a Mixture of Twenty Probes. One μ g of polyA⁺ RNA isolated from rat olfactory epithelium, brain, or spleen was size-fractionated in formaldehyde agarose, blotted onto a nylon membrane, and hybridized with a ³²P-labeled mixture of segments of 20 cDNA clones. The DNA segments were obtained by PCR using primers homologous to transmembrane domains 2 and 7.

Figure 4. The Protein Sequences Encoded by Ten Divergent cDNA Clones. Ten divergent cDNA clones were subjected to DNA sequence analyses and the protein sequence encoded by each was determined. Amino acid residues which are conserved in 60% or more of the proteins are shaded. The presence of seven hydrophobic domains (I-VII), as well as short conserved motifs shared with other members of the superfamily, demonstrate that these proteins belong to the seven transmembrane domain protein superfamily. Motifs conserved among members of the superfamily and the family of olfactory proteins include the GN in TM1 (transmembrane domain 1), the central W of TM4, the Y near the C-terminal end of TM5, and the NP in TM7. In addition, the DRY motif C-terminal to TM3 is common to many members of the G-protein-coupled superfamily. However, all of the proteins shown here share sequence motifs not found in other members of this superfamily and are clearly members of a novel family of proteins.

Figure 5. Positions of Greatest Variability in the Olfactory Protein Family. In this diagram, the protein encoded by cDNA clone I15 is shown traversing the plasma membrane seven times with its N-terminus located extracellularly, and its C-terminus intracellularly. The

-8-

vertical cylinders delineate the seven putative α -helices spanning the membrane. Positions at which 60% or more of the 10 clones shown in Figure 4 share the same residue as I15 are shown as white balls. More variable residues are shown as black balls. The high degree of variability encountered in transmembrane domains III, IV, and V is evident in this schematic.

Figure 6. The Presence of Subfamilies in a Divergent Multigene Family. Partial nucleotide sequences and deduced protein sequences were obtained for 18 different cDNA clones. Transmembrane domain V along with the flanking loop sequences, including the entire cytoplasmic loop between transmembrane domains V and VI, are shown here for each protein. Amino acid residues found in 60% or more of the clones in a given position are shaded (A). This region of the olfactory proteins (particularly transmembrane domain V) appears to be highly variable (see Figure 4). These proteins, however, can be grouped into subfamilies (B,C,D) in which the individual subfamily members share considerable homology in this divergent region of the protein.

Figure 7. Southern Blot Analyses with Non-crosshybridizing Fragments of Divergent cDNAs. Five μ g of rat liver DNA was digested with Eco RI (A) or Hind III (B), electrophoresed in 0.75% agarose, blotted onto a nylon membrane, and hybridized to the 32 P-labeled probes indicated. The probes used were PCR-generated fragments of: 1, clone F9 (identical to F12 in Figure 4); 2, F5; 3, F6; 4, I3; 5, I7; 6, I14; or 7, I15. The lane labeled "1-7" was hybridized to a mixture of the seven probes. The probes used showed either no crosshybridization or only trace crosshybridization with one another. The size markers on the left correspond to the four blots on the left (1-4) whereas the marker positions noted on the right correspond to the four blots on the right

-9-

(5-7, "1-7").

- Figure 8. Northern Blot Analysis with a Mix of Seven Divergent Clones. One μ g of polyA+ RNA from each of the tissues shown was size-fractionated, blotted onto a nylon membrane, and hybridized with a 32 P-labeled mixture of segments of seven divergent cDNA clones (see Legend to Figure 7).
- Figure 9. The amino acid and nucleic acid sequence of clone F3.
- Figure 10. The amino acid and nucleic acid sequence of clone F5.
- Figure 11. The amino acid and nucleic acid sequence of clone F6.
- Figure 12. The amino acid and nucleic acid sequence of clone F12.
- Figure 13. The amino acid and nucleic acid sequence of clone I3.
- Figure 14. The amino acid and nucleic acid sequence of clone I7.
- Figure 15. The amino acid and nucleic acid sequence of clone I8.
- Figure 16. The amino acid and nucleic acid sequence of clone I9.
- Figure 17. The amino acid and nucleic acid sequence of clone I14.

-10-

Figure 18. The amino acid and nucleic acid sequence of clone I15.

5 Figure 19. The amino acid and nucleic acid sequence of human clone H5.

10 Figure 20. The amino acid and nucleic acid sequence of clone J1, where the reading frame starts at nucleotide position 2.

Figure 21. The amino acid and nucleic acid sequence of clone J2.

15 Figure 22. The amino acid and nucleic acid sequence of clone J4, where the reading frame starts at nucleotide position 2.

20 Figure 23. The amino acid and nucleic acid sequence of clone J7, where the reading frame starts at nucleotide position 2.

25 Figure 24. The amino acid and nucleic acid sequence of clone J8, where the reading frame starts at nucleotide position 2.

Figure 25. The amino acid and nucleic acid sequence of clone J11.

30 Figure 26. The amino acid and nucleic acid sequence of clone J14, where the reading frame starts at nucleotide position 2.

35 Figure 27. The amino acid and nucleic acid sequence of clone J15, where the reading frame starts at nucleotide position 2.

-11-

Figure 28. The amino acid and nucleic acid sequence of clone J16, where the reading frame starts at nucleotide position 2.

5 Figure 29. The amino acid and nucleic acid sequence of clone J17, where the reading frame starts at nucleotide position 2.

10 Figure 30. The amino acid and nucleic acid sequence of clone J19, where the reading frame starts at nucleotide position 2.

15 Figure 31. The amino acid and nucleic acid sequence of clone J20, where the reading frame starts at nucleotide position 2.

20 Figure 32. SOUTHERN BLOT: Five micrograms of DNA isolated from 1. Human placenta, 2. NCI-H-1011 neuroblastoma cells, or 3. CHP 134 neuroblastoma cells were treated with the restriction enzyme A. Eco RI, B. Hind III, C. Bam HI, or D. Pst I, and then electrophoresed on an agarose gel and blotted onto a nylon membrane. The blotted DNA was hybridized to the ³²P-labeled H3/H5 sequence. An autoradiograph of the hybridized blot is shown with the
25 sizes of co-electrophoresed size markers noted in kilobases.

-12-

Detailed Description of the Invention

The invention provides an isolated nucleic acid, e.g. a DNA or cDNA molecule, encoding an odorant receptor. Such a
5 receptor is a receptor which binds an odorant ligand and include but not limited to pheromone receptors. An odorant ligand may include, but is not limited to, molecules which interact with the olfactory sensory neuron, molecules which interact with the olfactory cilia, pheromones, and molecules
10 which interact with structures within the vomeronasal organ.

The invention specifically provides the isolated cDNAs encoding odorant receptors the sequences of which are shown in Figures 9-31. The nucleic acid is most typically a cDNA
15 and encodes an insect, a vertebrate, a fish or a mammalian odorant receptor. The mammalian odorant receptor is preferably a human, rat, mouse or dog receptor. In an embodiment, human odorant receptor cDNA sequence and the correspondent protein is isolated (Figure 19).

20 In another embodiment, pheromone receptors are isolated and shown as clones J1, J2, J4, J7, J8, J11, J14, J15, J16, J17, J19 and J20 (Figures 20-31).

25 The invention further provides expression vectors containing cDNA which encodes odorant receptors. Such expression vectors are well known in the art and include in addition to the nucleic acid the elements necessary for replication and expression in a suitable hosts. Suitable hosts are well
30 known in the art and include without limitation bacterial hosts such as E. coli, animal hosts such as CHO cells, insect cells, yeast cells and like.

The invention also provides purified proteins encoding
35 odorant receptors. Such proteins may be prepared by

-13-

expression of the forementioned expression vectors in suitable host cells and recovery and purification of the receptors using methods well known in the art. Examples of such proteins include those having the amino acid sequences shown in figures 9-31.

The purified protein typically encodes an insect, vertebrate, fish or mammalian odorant receptor. The mammalian odorant receptor may be a human, rat, mouse or dog.

In one embodiment the invention provides a novel purified protein which belong to a class of proteins which have 7 transmembrane regions and a third cytoplasmic loop from the N-terminus which is approximately 17 amino acid long and to nucleic acid molecules encoding such proteins.

The invention provides methods of transforming cells which comprises transfecting a suitable host cell with a suitable expression vector containing nucleic acid encoding of the odorant receptor. Techniques for carrying out such transformations on cells are well known to those skilled in the art. (41,42) Additionally, the resulting transformed cells are also provided by the invention. These transformed cells may be either olfactory cells or non-olfactory cells. One advantage of using transformed non-olfactory cells is that the desired odorant receptor will be the only odorant receptor expressed on the cell's surface.

In order to obtain cell lines that express a single receptor type, standard procedures may be used to clone individual cDNAs or genes into expression vectors and then transfect the cloned sequences into mammalian cell lines. This approach has been used with sequences encoding some other members of the seven transmembrane domain superfamily

-14-

including the 5HT1c serotonin receptor. (43) The cited work illustrates how members of this superfamily transferred into cell lines may generate immortal cell lines that express high levels of the transfected receptor on the cell surface where it will bind ligand and that such abnormally expressed receptor molecules can transduce signals upon binding to ligand.

The invention also provides a method of identifying a desired odorant ligand which comprises contacting transformed non-olfactory cells expressing a known odorant receptor with a series of odorant ligands to determining which ligands bind to the receptors present on the non-olfactory cells.

Additionally, the invention provides a method of identifying a desired odorant receptor comprising contacting a series of transformed non-olfactory cells with a known odorant ligand and determining which odorant receptor binds with the odorant ligand.

The invention provides a method of detecting an odor which comprises: a) identifying a odorant receptor which binds the desired odorant ligand and; b) imbedding the receptor in a membrane such that when the odorant ligand binds to the receptor so identified a detectable signal is produced. In one embodiment of the invention the membrane used in this method is cellular, including a membrane of an olfactory cell or a synthetic membrane.

The ligand tested for may be the vapors emanating from Cocaine, Marijuana, Heroin, Hashish, Angel Dust, gasoline, decayed human flesh, alcohol, gun powder explosives, plastic explosives or firearms. In another embodiment the ligand tested for may be natural gas, a ph romone, toxic fumes,

-15-

noxious fumes or dangerous fumes.

In one embodiment of the invention the detectable signal is a lightbulb lighting up, a buzzer buzzing, a bell ringing,
5 a color change, phosphorescence, or radioactivity.

The invention further provides a method of quantifying the amount of an odorant ligand present in a sample which comprises utilizing the above-mentioned method for odor
10 detection and then quantifying the amount of signal produced.

The invention further provides a method of developing fragrances which comprises identifying a desired odorant
15 receptor by the above method, then contacting non-olfactory cells, which have been transfected with an expression vector containing nucleic acid encoding the desired odorant receptor such that the receptor is expressed upon the surface of the non-olfactory cell, with a series of
20 compounds to determine which compound or compounds bind the receptor.

The invention provides to a method of identifying an "odorant fingerprint" which comprises contacting a series of
25 cells, which have been transformed such that each express a known odorant receptor, with a desired sample and determining the type and quantity of the odorant ligands present in the sample.

30 The invention provides a method of identifying odorant ligands which inhibit the activity of a desired odorant receptor which comprises contacting the desired odorant receptor with a series of compounds and determining which compounds inhibit the odorant ligand - odorant receptor
35 interaction.

-16-

The invention also provides for a method of identifying appetite suppressant compounds which comprises identifying odorant ligands by the method mentioned in the preceding paragraph wherein the desired odorant receptor is that which is associated with the perception of food. Additionally, the invention provides a method of controlling appetite in a subject which comprises contacting the olfactory epithelium of the subject with these odorant ligands. Further the invention provides a nasal spray, to control appetite comprising the compounds identified by the above method in a suitable carrier.

The invention provides a method of trapping odors which comprises contacting a membrane which contains multiples of the desired odorant receptor, with a sample such that the desired odorant ligand is absorbed by the binding of the odorant ligand to the odorant receptor. The invention also provides an odor trap employing this method.

The invention also provides a method of controlling pest populations which comprises identifying odorant ligands by the method mentioned above which are alarm odorant ligands and spraying the desired area with the identified odorant ligands. Additionally, provided by the invention is a method of controlling a pest population which comprises identifying odorant ligands by the above mentioned method, which interfere with the interaction between the odorant ligands and the odorant receptors which are associated with fertility. In one embodiment the pest population is a population of insects or rodents, including mice and rats.

The invention also provides a method of promoting fertility which comprises identifying odorant ligands which interact with the odorant receptors associated with fertility by the above mentioned method. Further, the invention provides a

-17-

method of inhibiting fertility which comprises employing the above mentioned method to identifying odorant ligands which inhibit the interaction between the odorant ligands and the odorant receptors associated with fertility.

5

This invention is illustrated in the Experimental Detail section which follow. These sections are set forth to aid in an understanding of the invention but are not intended to, and should not be construed to, limit in any way the invention as set forth in the claims which follow thereafter.

10

EXPERIMENTAL DETAILS

15

MATERIALS AND METHODS

Polymerase Chain Reaction

RNA was prepared from the olfactory epithelia of Sprague Dawley rats according to Chirgwin et al. (40) or using RNazol B (Cinna/Biotech) and then treated with DNase I (0.1 unit/ μ g RNA) (Promega). In order to obtain cDNA, this RNA was incubated at 0.1 μ g/ μ l with 5 μ M random hexamers (Pharmacia) 1 mM each of dATP, dCTP, dGTP, TTP, and 2 units/ μ l RNase inhibitor (Promega) in 10 mM TrisCl (pH 8.3), 50 mM KCl, 2.5 mM MgCl₂, and 0.001% gelatin for 10 min. at 22°C, and then for a further 45 min. at 37°C following the addition of 20 u./ μ l of Moloney murine leukemia virus reverse transcriptase (BRL). After heating at 95°C for 3 min., cDNA prepared from 0.2 μ g of RNA was used in each of a series of polymerase chain reactions (PCR) containing 10 mM TrisCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.001% gelatin, 200 μ M each of dATP, dCTP, dGTP, and TTP, 2.5 u. Taq polymerase (Perkin Elmer Cetus), and 2 μ M of each PCR

35

-18-

primer. PCR reactions were performed according to the following schedule: 96°C for 45 sec., 55°C for 4 min. (or 45°C for 2 min.), 72°C for 3 min. with 6 sec. extension per cycle for 48 cycles. The primers used for PCR were a series of degenerate oligonucleotides made according to the amino acid sequences found in transmembrane domain 2 and 7 of a variety of different members of the 7 transmembrane domain protein superfamily (19). The regions used correspond to amino acids number 60-70 and 286-295 of clone I15 (Figure 4). Each of five different 5' primers were used in PCR reactions with each of six different 3' primers. The 5' primers had the sequences:

15 C AC A C CT
A1, AATTGGATICTIGTIAATCTIGCIGTIGCIGCIGA;

 C C CA A C C
A2, AATTATTTTCTIGTIAATCTIGCITTIGCIGA;

20 CCA CC A C
A3, AATTTITTTATATITCICTIGCITGIGCIGA;

 A T C T ACT C
A4, CGITTICTIATGTGTAACCTITGCTTTGCIGA;

25 C CT TG
A5, ACIGTITATATIACICATCTIACIATIGCIGA.

The 3' primers were:

30 TTA T CAG C C A
B1, CTGICGGTTCATIAAIAACATAIATATIGGGTT;

 TG GA G G A A
35 B2, GATCGTTIAGACAACAATAIATATIGGGTT;

 A G G A
B3, TCIATGTTAAAGTIGTATAIATATIGGGTT;

40 T G G A A
B4, GCCTTIGTAAAIATIGCATAIAGGAAIGGGTT;

 G AGA G G G A
45 B5, AAATCIGGGCTICGICAATAIATCAIIGGGTT;

 CT CT G G G G A

-19-

B6, GAIGAICCIACAAAAAATAIATAAAIGGGTT.

5 An aliquot of each PCR reaction was analyzed by agarose gel electrophoresis and bands of interest were amplified further by performing PCR reactions on pipet tip (approx. 1 μ l) plugs of the agarose gels containing those DNAs. Aliquots of these semi-purified PCR products were digested with the restriction enzymes Hae III or Hinf I and the digestion
10 products were compared with the undigested DNAs on agarose gels.

Isolation and Analysis of cDNA Clones

15 CDNA libraries were prepared according to standard procedures (41, 42) in the cloning vector, λ ZAP II (Stratagene) using poly A⁺ RNA prepared from Sprague Dawley rat epithelia (see above) or from an enriched population of olfactory neurons which had been obtained by a 'panning'
20 procedure, using an antibody against the H blood group antigen (Chembiomed) found on a large percentage of rat olfactory neurons. In initial library screens, 8.5×10^5 independent clones from the olfactory neuron library and 1.8×10^6 clones from the olfactory epithelium library were
25 screened (41) with a ³²P-labeled probe (prime-it, Stratagene) consisting of a pool of gel-isolated PCR products obtained using primers A4 and B6 (see above) in PCR reactions using as template, olfactory epithelium cDNA, rat liver DNA, or DNA prepared from the two cDNA libraries. In
30 later library screens, a mixture of PCR products obtained from 20 cDNA clones with the A4 and B6 primers was used as probe ('P1' probe). In initial screens, phage clones were analyzed by PCR using primers A4 and B6 and those which showed the appropriate size species were purified. In later
35 screens, all position clones were purified, but only those that could be amplified with the B6 primer and a primer

-20-

specific for vector sequence were analyzed further. To obtain plasmids from the isolated phage clones, phagemid rescue was performed according to the instructions of the manufacturer of λ ZAP II (Stratagene). DNA sequence analysis was performed on plasmid DNAs using the Sequenase system (USB), initially with the A4 and B6 primers and later with oligonucleotide primers made according to sequences already obtained.

10 Northern and Southern Blot Analyses

For Northern blots, poly A⁺ RNAs from various tissues were prepared as described above or purchased from Clontech. One μ g of each RNA was size fractionated on formaldehyde agarose gels and blotted onto nylon membranes (41, 42). For Southern blots, genomic DNA prepared from Sprague Dawley rat liver was digested with the restriction enzymes Eco RI or Hind III, size fractionated on agarose gels and blotted onto nylon membranes (41, 42). The membranes were dried at 80°C, and then prehybridized in 0.5 M sodium phosphate buffer (pH 7.3) containing 1% bovine serum albumin and 4% sodium dodecyl sulfate. Hybridization was carried out in the same buffer at 65°-70°C for 14-20 hrs. with DNAs labeled with ³²P. For the first Northern blot shown, the 'P1' probe (see above under cDNA clone isolation) was used. For the second Northern blot shown, a mix of PCR fragments from seven divergent cDNA clones was used. For Southern blots, the region indicated in clone I15 by amino acids 118 through 251 was amplified from a series of divergent cDNA clones using PCR. The primers used for these reactions had the sequences:

P1, ATGGCITATGATCGITATGTIGC, and

35 P4, AAIAGIGAIACIATIGAIAGATGIGAICC

-21-

These DNAs (or a DNA encompassing transmembrane domains 2 through 7 for clone F6) were labeled and tested for crosshybridization at 70°C. Those DNAs which did not show appreciable crosshybridization were hybridized individually, or as a pool to Southern blots at 70°C.

Rat Sequences used to obtain similar sequences expressed in Humans

There are genes similar to the rat genes discussed above present in humans, these genes may be readily isolated by screening human gene libraries with the cloned rat sequences or by performing PCR experiments on human genomic DNA with primers homologous to the rat sequences. First, PCR experiments were performed with genomic DNA from rat, human, mouse, and several other species. When primers homologous to transmembrane domains 2 and 6 (the A4/B6 primer set used to isolate the original rat sequences) were used, DNA of the appropriate size was amplified from rat, human and mouse DNAs. When these primary PCR reactions were subsequently diluted and subjected to PCR using primers to internal sequences (P1 and P4 primers), smaller DNA species were amplified whose size was that seen when the same primers were used in PCR reactions with the cloned rat cDNAs. Similarly, when the secondary PCR was performed with one outer primer together with one inner primer (ie. A4/P4 or P1/B6), amplified DNAs were obtained whose sizes were also consistent with the amplification of genes similar in sequence and organization to the cloned rat cDNAs. Second, a mix of segments from 20 of the rat cDNAs ('P1" probe) was used to screen libraries constructed from human genomic DNAs. Hybridization under high or low stringency conditions reveals the presence of a large number of cloned human DNA segments that are homologous to the rat sequences. Finally, RNA from a human olfactory tumor (neuroesthesioma,

-22-

NCI-H-1011) cell line has been examined for sequences homologous to those cloned in the rat. cDNA prepared from this RNA was subjected to PCR with the A4/B6 primer set and a DNA species of the appropriate size was seen. This DNA
5 was subcloned and partially sequenced and clearly encodes a member of the olfactory protein family identified in the rat.

The inserted sequence in human clones H3/H5 was amplified by
10 PCR with the A4/B6 primers, gel purified, and then labeled with ³²P. The labeled DNA was then hybridized to restriction enzyme human placenta. Multiple hybridizing species were observed with each DNA (See Figure 32). This observation is consistent with the presence of a family of
15 odorant receptor genes in the human genome.

The sequence of clone H5 is hereby shown in Figure 19. In addition, the translated protein sequence is shown in Figure
20 19.

In order to identify odorant receptors in other species, degenerated primer oligonucleotides homologous to conserved regions within the rat odorant receptor family may be used in PCR reactions with genomic DNA or with cDNA prepared from
25 olfactory tissue RNA from those species.

RESULTS

Cloning the Gene Family

30 A series of degenerate oligonucleotides were designated which could anneal to conserved regions of members of the superfamily of G-protein coupled seven transmembrane domain receptor genes. Five degenerate oligonucleotides (A1-5; see Experimental Procedures) matching sequences within
35 transmembrane domain 2, and six degenerate oligonucleotides

-23-

(B1-6) matching transmembrane domain 7 were used in all combinations in PCR reactions to amplify homologous sequences in cDNA prepared from rat olfactory epithelium RNA. The amplification products of each PCR reaction were then analyzed by agarose gel electrophoresis. Multiple bands were observed with each of the primer combinations. The PCR products within the size range expected for this family of receptors (600 to 1300 bp) were subsequently picked and amplified further with the appropriate primer pair in order to isolate individual PCR bands. Sixty-four PCR bands isolated in this fashion revealed only one or a small number of bands upon agarose gel electrophoresis. Representatives of these isolated PCR products are shown in Figure 2A.

The isolated PCR products were digested with the endonuclease, Hae III or Hinf I, which recognize four base restriction sites and cut DNA at frequent intervals. In most instances, digestion of the PCR product with Hinf I generated a set of fragments whose molecular weights sum to the size of the original DNA (Figure 2B). These PCR bands are therefore likely to each contain a single DNA species. In some cases, however, restriction digestion yielded a series of fragments whose molecular weights sum to a value greater than that of the original PCR product. The most dramatic example is shown in Figure 2 where the 710 bp, PCR 13 DNA, is cleaved by Hinf I to yield a very large number of restriction fragments whose sizes sum to a value five- to ten-fold greater than that of the original PCR product. These observations indicated that PCR product 13 consists of a number of different species of DNA, each of which could be amplified with the same pair of primer oligonucleotides. In addition, when PCR experiments similar to those described were performed using cDNA library DNAs as templates, a 710 bp PCR product was obtained with the PCR13 primer pair

-24-

(A4/B6) with DNA from olfactory cDNA libraries, but not a glioma cDNA library. Moreover, digestion of one of this 710 bp product also revealed the presence of multiple DNA species. In other cases (see PCR product 20, for example),
5 digestion yielded a series of restriction fragments whose molecular weights also sum to a size greater than the starting material. Further analysis, however, revealed that the original PCR product consisted of multiple bands of similar but different sizes.

10

In order to determine whether the multiple DNA species present in PCR 13 encode members of a family of seven transmembrane domain proteins, PCR 13 DNA was cloned into the plasmid vector Bluescript and five individual clones
15 were subjected to DNA sequence analysis. Each of the five clones exhibited a different DNA sequence, but each encoded a protein which displayed conserved features of the superfamily of seven transmembrane domain receptor proteins. In addition, the proteins encoded by all five clones shared
20 distinctive sequence motifs not found in other superfamily members indicating they were all members of a new family of receptors.

To obtain full-length cDNA clones, cDNA libraries prepared
25 from olfactory epithelium RNA or from RNA of an enriched population of olfactory sensory neurons were screened. The probe used in these initial screens was a mixture of PCR 13 DNA as well as DNA obtained by amplification of rat genomic DNA or DNA from two olfactory cDNA libraries with the same
30 primers used to generate PCR 13 (A4 and B6 primers). Hybridizing plaques were subjected to PCR amplification with the A4/B6 primer set and only those giving a PCR product of the appropriate size (approximately 710 bp) were purified. The frequency of such positive clones in the enriched
35 olfactory neuron cDNA library was approximately five times

-25-

greater than the frequency in the olfactory epithelium cDNA library. The increased frequency of positive clones observed in the olfactory neuron library is comparable to the enrichment in olfactory neurons generally obtained in the purification procedure.

The original pair of primers used to amplify PCR 13 DNA were then used to amplify coding segments of 20 different cDNA clones. A mix of these PCR products were labeled and used as probe for further cDNA library screens. This mixed probe was also used in a Northern blot (Figure 3) to determine whether the expression of the gene family is restricted to the olfactory epithelium. The mixed probe detects two diffuse bands centered at 2 and 5 kb in RNA from olfactory epithelium; no hybridization can be detected in brain or spleen. (Later experiments which examined a larger number of tissue RNAs with a more restricted probe will be shown below.) Taken together, these data indicate the discovery of a novel multigene family encoding seven transmembrane domain proteins which are expressed in olfactory epithelium, and could be expressed predominantly or exclusively in olfactory neurons.

The Protein Sequences of Numerous, Olfactory-specific Members of the Seven Transmembrane Domain Superfamily

Numerous clones were obtained upon screening cDNA libraries constructed from olfactory epithelium and olfactory neuron RNA at high stringency. Partial DNA sequences were obtained from 36 clones; 18 of these cDNA clones are different, but all of them encode proteins which exhibit shared sequence motifs indicating that they are members of the family identified in PCR 13 DNA. A complete nucleotide sequence was determined for coding regions of ten of the most divergent clones (Figure 4). The deduced protein sequences

-26-

of these cDNAs defines a new multigene family which shares sequence and structural properties with the superfamily of neurotransmitter and hormone receptors that traverse the membrane seven times. This novel family, however, exhibits
5 features different from any other member of the receptor superfamily thus far identified.

Each of the ten sequences contains seven hydrophobic stretches (19-26 amino acids) that represent potential
10 transmembrane domains. These domains constitute the regions of maximal sequence similarity to other members of the seven transmembrane domain superfamily (see legend to Figure 4). On the basis of structural homologies with rhodopsin and the β -adrenergic receptors, (19) it is likely that the amino
15 termini of the olfactory proteins are located on the extracellular side of the plasma membrane and the carboxyl termini are located in the cytoplasm. In this scheme, three extracellular loops alternate with three intracellular loops to link the seven transmembrane domains (see Figure 5).
20 Analysis of the sequences in figure 4 demonstrates that the olfactory proteins, like other members of the receptor superfamily, display no evidence of an N-terminal signal sequence. As in several other superfamily members, a potential N-linked glycosylation site is present in all ten
25 proteins within the short N-terminal extracellular segment. Other structural features conserved with previously identified members of the superfamily included cysteine residues at fixed positions within the first and second extracellular loops that are thought to form a disulfide
30 bond. Finally, many of the olfactory proteins reveal a conserved cysteine within the C-terminal domain which may serve as a palmitoylation site anchoring this domain to the membrane (21). These features, taken together with several short, conserved sequence motifs (see legend to Figure 4),
35 clearly define this new family as a member of the

-27-

superfamily of genes encoding the seven transmembrane domain receptors.

There are, however, important differences between the
5 olfactory protein family and the other seven transmembrane
domain proteins described previously and these differences
may be relevant to proposed function of these proteins in
odor recognition. Structure-function experiments involving
10 in vitro mutagenesis suggest that adrenergic ligands
interact with this class of receptor molecule by binding
within the plane of the membrane (22, 20). Not
surprisingly, small receptor families that bind the same
class of ligands, such as the adrenergic and muscarinic
acetylcholine receptor families exhibit maximum sequence
15 conservation (often over 80%) within the transmembrane
domains. In contrast, the family of receptors discussed in
this application shows striking divergence within the third,
fourth, and fifth transmembrane domains (Figure 4). The
variability in the three central transmembrane domains is
20 highlighted schematically in Figure 5. The divergence in
potential ligand binding domains is consistent with the idea
that the family of molecules cloned is capable of
associating with a large number of odorant of diverse
molecular structure.

25 Receptors which belong to the superfamily of seven
transmembrane domain proteins interact with G-proteins to
generate intracellular signals. In vitro mutagenesis
experiments indicate that one site of association between
30 receptor and G-protein resides within the third cytoplasmic
loop (22, 23). The sequence of this cytoplasmic loop in 18
different clones we have characterized is shown in Figure
6A. This loop which is often quite long and of variable
length in the receptor superfamily is relatively short (only
35 17 amino acids) and of fixed length in the 18 clones

-28-

examined. Eleven of the 18 different clones exhibit the sequence motif K/R I V S S I (or a close relative) at the N-terminus of this loop. Two of the cDNA clones reveal a different H I T C/W A V motif at this site. If this short
5 loop is a site of contact with G-proteins, it is possible that the conserved motifs may reflect sites of interaction with different G-proteins that activate different intracellular signalling systems in response to odors. In addition, the receptors cloned reveal several serine or
10 threonine residues within the third cytoplasmic loop. By analogy with other G-protein coupled receptors, these residues may represent sites of phosphorylation for specific receptor kinases involved in desensitization. (24)

15 Subfamilies within the Multigene Family

Figure 6A displays the sequences of the fifth transmembrane domain and the adjacent cytoplasmic loop encoded by L8 of the cDNA clones we have analyzed. As a group, the 18
20 sequences exhibit considerable divergence within this region. The multigene family, however, can be divided into subfamilies such that the members of a given subfamily share significant sequence conservation. In subfamily B, clones F12 and F13, for example, differ from one another at only
25 four of 44 positions (91% identify), and clearly define a subfamily. Clones F5 and I11 (subfamily D) differ from F12 and F13 at 34-36 positions within this region and clearly define a separate subfamily. Thus, this olfactory-specific multigene family consists of highly divergent subfamilies.
30 If these genes encode odor receptors, it is possible that members of the divergent subfamilies bind odorant of widely differing structural classes. Members of the individual subfamilies could therefore recognize more subtle differences between molecules which belong to the same
35 structural class of molecules structures.

-29-

The Size of the Multigene Family

Genomic Southern blotting experiments were performed and genomic libraries were screened to obtain an estimate of the sizes of the multigene family and the member subfamilies encoding the putative odor receptors. DNAs extending from the 3' end of transmembrane domain 3 to the middle of transmembrane domain 6 were synthesized by PCR from DNA of seven of the divergent cDNA clones (Figure 4). In initial experiments, these DNAs were labeled and hybridized to each other to define conditions under which minimal crosshybridization would be observed among the individual clones. At 70°C, the seven DNAs showed no crosshybridization, or crosshybridized only very slightly. The trace levels of crosshybridization observed are not likely to be apparent upon genomic Southern blot analysis where the amounts of DNA are far lower than in the test cross.

Probes derived from these seven DNAs were annealed under stringent conditions, either individually or as a group, to Southern blots of rat liver DNA digested with the restriction endonucleases Eco RI or Hind III (Figure 7). Examination of the Southern blots reveals that all but one of the cDNAs detects a relatively large, distinctive array of bands in genomic DNA. Clone I15 (probe 7), for example, detects about 17 bands with each restriction endonuclease, whereas clone F9 (probe 1) detects only about 5-7 bands with each enzyme. A single band is obtained with clone I7 (probe 5). PCR experiments using nested primers (TM2/TM7 primers followed by primers to internal sequences) and genomic DNA as template indicate that the coding regions of the members of this multigene family, like those of many members of the G-protein coupled superfamily, may not be interrupted by introns. This observation, together with the fact that most

-30-

of the probes only encompasses 400 nucleotides suggests that each band observed in these experiments is likely to represent a different gene. These data suggest that the individual probes chosen are representatives of subfamilies which range in size from a single member to as many as 17 members. A total of about 70 individual bands were detected in this analysis which could represent the presence of at least 70 different genes. Although the DNA probes used in these blots did not crosshybridize appreciably with each other, it is possible that a given gene might hybridize to more than one probe, resulting in an overestimate of gene number. However, it is probable that the total number of bands only reflects a minimal estimate of gene number since it is unlikely that we have isolated representative cDNAs from all of the potential subfamilies and the hybridizations were performed under conditions of very high stringency.

A more accurate estimate of the size of the olfactory-specific gene family was obtained by screening rat genomic libraries. The mix of the seven divergent probes used in Southern blots, or the mix of 20 different probes used in our initial Northern blots (see Figure 3), were used as hybridization probes under high (65°C) or lowered (55°C) stringency conditions in these experiments. Nested PCR (see above) was used to verify that the clones giving a positive signal under low stringency annealing conditions were indeed members of this gene family. It is estimated from these studies that there are between 100 and 200 positive clones per haploid genome. The estimate of the size of the family obtain from screens of genomic libraries again represents a lower limit. Given the size of the multigene family, one might anticipate that many of these genes are linked such that a given genomic clone may contain multiple genes. Thus the data from Southern blotting and screens of genomic libraries indicate that the multigene family identified

-31-

consists of one to several hundred member genes which can be divided into multiple subfamilies.

It should be noted that the cDNA probes isolated may not be representative of the full complement of subfamilies within the larger family of olfactory proteins. The isolation of cDNAs, for example, relies heavily on PCR with primers from transmembrane domains 2 and 7 and biases our clones for homology within these regions. Thus, estimates of gene number as well as subsequent estimates of RNA abundance should be considered as minimal.

Expression of the Members of this Multigene Family

Additional Northern blot analyses were performed to demonstrate that expression of the members of this gene family is restricted to the olfactory epithelium. (Figure 8) Northern blot analysis with a mixed probe consisting of the seven divergent cDNAs used above reveals two diffuse bands about 5 and 2 kb in length in olfactory epithelium RNA. This pattern is the same as that seen previously with the mix of 20 DNAs. No annealing is observed to RNA from the brain or retina or other, nonneural tissues, including lung, liver, spleen, and kidney.

An estimate of the level of expression of this family can be obtained from screens of cDNA libraries. The frequency of positive clones in cDNA libraries made from olfactory epithelium RNA suggests that the abundance of the RNAs in the epithelium is about one in 20,000. The frequency of positive clones is approximately five-fold higher in a cDNA library prepared from RNA from purified olfactory neurons (in which 75% of the cells are olfactory neurons). The increased frequency of positive clones obtained in the olfactory neuron cDNA library is comparable to the

-32-

enrichment we obtain upon purification of olfactory neurons. These observations suggest that this multigene family is expressed largely, if not solely, in olfactory neurons and may not be expressed in other cell types within the epithelium. If each olfactory neuron contains 10^5 mRNA molecules, from the frequency of positive clones we predict that each neuron contains only 25-30 transcripts derived from this gene family. Since the family of olfactory proteins consists of a minimum of a hundred genes, a given olfactory neuron could maximally express only a proportion of the many different family members. These values thus suggest that olfactory neurons will exhibit significant diversity at the level of expression of these olfactory proteins.

15 Identification of pheromone receptors in vomeronasal organ
The vomeronasal organ (vomeronasal gland) is an accessory olfactory structure that is located near the nasal cavity. Like the olfactory epithelium of the nasal cavity, the
20 olfactory epithelium of the vomeronasal organ contains olfactory sensory neurons. The vomeronasal organ is believed to play an important role in the sensing of pheromones in numerous species. Pheromones are believed to have profound effects on both physiological and behavioral
25 aspects of reproduction. the identification of pheromone receptors would permit the identification of the pheromones themselves. It would also enable one to identify agonists or antagonists that would either mimic the pheromones or block the pheromone receptors from transducing pheromone signals.
30 Such information would be important to the development of species specific pesticides and, conversely, to animal husbandry. The identification of pheromone receptors in human could ultimately lead to the development of contraceptives or to treatments for infertility in humans.
35 It is likely that the identification of pheromone receptors

-33-

in low mammals such as rodents would lead to the identification of similar receptors in human.

5 In order to identify potential pheromone receptors, we isolate RNA from the vomeronasal organs of female rats and prepared cDNA from this RNA. The cDNA was subjected to PCR with several different pairs of degenerate oligonucleotide primers that match sequences present in the rat odorant receptor family. The PCR products were subcloned and the
10 nucleotide sequences of the subcloned DNAs were determined. Each of the subcloned DNAs encodes a protein that belongs to the odorant receptor family. The sequences of the following vomeronasal subclones are shown: J1, J2, J4, J7, J8, J11, J14, J15, J16, J17, J19, J20. In a few cases (J2, J4), the
15 same sequence was amplified with two different primer pairs and the sequence shown is a composite of the two sequences. It is possible that one or more of these molecules, or closely related molecules, serve as pheromone receptors in the rat.

20

DISCUSSION

The mammalian olfactory system can recognize and discriminate a large number of odorous molecules.
25 Perception in this system, as in other sensory systems, initially involves the recognition of external stimuli by primary sensory neurons. This sensory information is then transmitted to the brain where it is decoded to permit the discrimination of different odors. Elucidation of the logic
30 underlying olfactory perception is likely to require the identification of the specific odorant receptors, the analysis of the extent of receptor diversity and receptor specificity, as well as an understanding of the pattern of receptor expression in the olfactory epithelium.

35

-34-

The odorant receptors are thought to transduce intracellular signals by interacting with G-proteins which activate second messenger systems (12, 13, 14, 15). These proteins are clearly members of the family of G-protein coupled receptors which traverse the membrane seven times (19). The odorant receptors should be expressed specifically in the tissue in which odorant are recognized. The family of olfactory proteins cloned is expressed in the olfactory epithelium. Hybridizing RNA is not detected in brain or retina, or in a host of nonneural tissues. Moreover, expression of this gene family the epithelium may be restricted to olfactory neurons. The family of odorant receptors must be capable of interacting with extremely diverse molecular structures. The genes cloned are members of any extremely large multigene family which exhibit variability in regions thought to be important in ligand binding. The possibility that each member of this large family of seven transmembrane proteins is capable of interacting with only one or a small number of odorant provides a plausible mechanism to accommodate the diversity of odor perception. The properties of the gene family identified suggests that this family is likely to encode a large number of distinct odorant receptors.

25 Size of the Multigene Family

The size of the receptor repertoire is likely to reflect the range of detectable odors and the degree of structural specificity exhibited by the individual receptors. It has been estimated that humans can identify over 10,000 structurally-distinct odorous ligands. However, this does not necessarily imply that humans possess an equally large repertoire of odorant receptors. For example, binding studies in lower vertebrates suggest that structurally-related odorant may activate the same receptor molecules.

-35-

In fish which smell amino acids, the binding of alanine to isolated cilia can be competed by other small polar residues (threonine and serine), but not by the basic amino acids, lysine or arginine (11). These data suggest that individual
5 receptors are capable of associating with several structurally-related ligands, albeit with different affinities. Stereochemical models of olfactory recognition in mammals (25) (based largely on psychophysical, rather than biophysical data) have suggested existence of several
10 primary odor groups including camphoraceous, musky, peppermint, ethereal, pungent, and putrid. In such a model, each group would contain odorant with common molecular configurations which bind to common receptors and share similar odor qualities.

15 Screens of genomic libraries with mixed probes consisting of divergent family members detect approximately 100 to 200 positive clones per genome. The present estimate of at least 100 genes provides only a lower limit since it is
20 likely that the probes used do not detect all of the possible subfamilies. Moreover, it is probable that many of these genes are linked such that a given genomic clone may contain multiple genes. It is therefore expected that the actual size of the gene family may be considerably higher
25 and this family of putative odorant receptors could constitute one of the largest gene families in the genome.

The characterization of a large multigene family encoding putative odorant receptors suggests that the olfactory
30 system utilizes a far greater number of receptors than the visual system. Color vision, for example, allows the discrimination of several hundred hues, but is accomplished by only three different photoreceptors (1, 2, 3 and 4). The photoreceptors each have different, but overlapping
35 absorption spectra which cover the entire spectrum of

-36-

visible wavelengths. Discrimination of color results from comparative processing of the information from these three classes of photoreceptors in the brain. Whereas three photoreceptors can absorb light across the entire visible spectrum, our data suggest that a small number of odorant receptors cannot recognize and discriminate the full spectrum of distinct molecular structures perceived by the mammalian olfactory system. Rather, olfactory perception probably employs an extremely large number of receptors each capable of recognizing a small number of odorous ligands.

Diversity within the Gene Family and the Specificity of Odor Recognition

The olfactory proteins identified in this application are clearly members of the superfamily of receptors which traverse the membrane seven times. Analysis of the proteins encoded by the 18 distinct cDNAs we have cloned reveals structural features which may render this family particularly well suited for the detection of a diverse array of structurally distinct odorants. Experiments with other members of this class of receptors suggest that the ligand binds to its receptor within the plane of the membrane such that the ligand contacts many, if not all of the transmembrane helices. The family of olfactory proteins can be divided into several different subfamilies which exhibit significant sequence divergence within the transmembrane domains. Nonconservative changes are commonly observed within blocks of residues in transmembrane regions 3, 4, and 5 (Figures 4, 5, 6); these blocks could reflect the sites of direct contact with odorous ligands. Some members, for example, have acidic residues in transmembrane domain 3, which in other families are thought to be essential for binding aminergic ligands (20) while other members maintain hydrophobic residues at these positions.

-37-

This divergence within transmembrane domains may reflect the fact that the members of the family of odorant receptors must associate with odorant of widely different molecular structures.

5

These observations suggest a model in which each of the individual subfamilies encode receptors which bind distinct structural classes of odorant. Within a given subfamily, however, the sequence differences are far less dramatic and are often restricted to a small number of residues. Thus, the members of a subfamily may recognize more subtle variations among odor molecules of a given structural class. At a practical level, individual subfamilies may recognize grossly different structures such that one subfamily may associate, for example, with the aromatic compound, benzene and its derivatives, whereas a second subfamily may recognize odorous, short chain, aliphatic molecules. Subtle variations in the structure of the receptors within, for example, the hypothetical benzene subfamily could facilitate the recognition and discrimination of various substituted derivatives such as toluene, xylene or phenol. It should be noted that such a model, unlike previous stereochemical models, does not necessarily predict that molecules with similar structures will have similar odors. The activation of distinct receptors with similar structures could elicit different odors, since perceived odor will depend upon higher order processing of primary sensory information.

30

Evolution of the Gene Family and the Generation of Diversity

Preliminary evidence from PCR analyses suggests that members of this family of olfactory proteins are conserved in lower vertebrates as well as invertebrates. This gene family presumably expanded over evolutionary time providing mammals with the ability to recognize an increasing diversity of

35

-38-

odorant. Examination of the sequences of the family members cloned from mammals provides some insight into the evolution of this multigene family. Although the chromosomal loci encoding these genes has yet to be characterized, it is likely that at least some member genes will be tandemly arranged in a large cluster as is observed with other large multigene families. A tandem array of this sort provides a template for recombination events including unequal crossing over and gene conversion, that can lead to expansion and further diversification of the sort apparent among the family members we have cloned (26).

The multigene family encoding the olfactory proteins is large: all of the member genes clearly have a common ancestral origin, but have undergone considerable divergence such that individual genes encode proteins that share from 40-80% amino acid identity. Subfamilies are apparent with groups of genes sharing greater homology among themselves than with members of other subfamilies. Examination of the sequences of even the most divergent subfamilies, however, reveals a pattern in which several blocks of conserved residues are interspersed with variable regions. This segmental homology is conceptually similar to the organization of framework and hypervariable domains within the families of immunoglobulin and T cell receptor variable region sequences (27, 28). This analogy goes beyond structural organization and may extend to the function of these two gene families: each family consists of a large number of genes which have diversified over evolutionary time to accommodate the binding of a highly diverse array of ligands. The evolutionary mechanisms responsible for the diversification and maintenance of these large gene families may also be similar. It has been suggested that gene conversion has played a major role in the evolution of immunoglobulin and T cell receptor variable domains (29, 30

-39-

- and 31). Analysis of the sequence of the putative olfactory receptors reveals at least one instance where a motif from a variable region of one subfamily is found imbedded in the otherwise divergent sequence of a second subfamily, suggesting that conversion has occurred. Such a mixing of motifs from one subfamily to another over evolutionary time would provide additional combinatorial possibilities leading to the generation of diversity.
- It should be noted, however, that the combinatorial joining of gene segments by DNA rearrangement during development, which is characteristic of immunoglobulin loci (27), is not a feature of the putative odor receptor gene family. No evidence for DNA rearrangement to generate the diversity of genes cloned has been observed. The entire coding region has been sequenced along with parts of the 5' and 3' untranslated regions of 10 different cDNA clones. The sequences of the coding regions are all different; no evidence has been obtained for constant regions that would suggest DNA rearrangement of the sort seen in the immune system. The observations indicate that the diversity olfactory proteins are coded by a large number of distinct gene sequences.
- Although it is unlikely from the data that DNA rearrangement is responsible for the generation of diversity among the putative odorant receptors, it remains possible that DNA rearrangements may be involved in the regulation of expression of this gene family. If each olfactory neuron expresses only one or a small number of genes, then a transcriptional control mechanism must be operative to choose which of the more than one hundred genes within the family will be expressed in a given neuron. Gene conversion from one of multiple silent loci into a single active locus, as observed for the trypanosome-variable surface

-40-

glycoproteins (32), provides one attractive model. The gene conversion event could be stochastic, such that a given neuron could randomly express any one of several hundred receptor genes, or regulated (perhaps by positional information), such that a given neuron could only express one or a small number of predetermined receptor types. Alternatively, it is possible that positional information in the olfactory epithelium controls the expression of the family of olfactory receptors by more classical mechanisms that do not involve DNA rearrangement. What ever mechanisms will regulate the expression of receptor genes within this large, multigene family, these mechanisms must accommodate the requirement that olfactory neurons are regenerated every 30-60 days (8) and therefore the expression of the entire repertoire of receptors must be accomplished many times during the life of an organism.

Receptor Diversity and the Central Processing of Olfactory Information

The results suggest the existence of a large family of distinct odorant receptors. Individual members of this receptor family are likely to be expressed by only a small set of the total number of olfactory neurons. The primary sensory neurons within the olfactory epithelium will therefore exhibit significant diversity at the level of receptor expression. The question then emerges as to whether neurons expressing the same receptors are localized in the olfactory epithelium. Does the olfactory system employ a topographic map to discriminate among the numerous odorant? The spatial organization of distinct classes of olfactory sensory neurons, as defined by receptor expression, can now be determined by using the procedures of in situ hybridization and immunohistochemistry with probes specific for the individual receptor subtypes. This

-41-

information should help to distinguish between different models that have been proposed to explain the coding of diverse odorant stimuli (33).

5 In one model, sensory neurons that express a given receptor and respond to a given odorant may be localized within defined positions within the olfactory epithelium. This topographic arrangement would also be reflected in the projection of olfactory sensory axons into discrete regions
10 (glomeruli) within the olfactory bulb. In this scheme, the central coding to permit the discrimination of discrete odorant would depend, in part, on the spatial segregation of different receptor populations. Attempts to discern the topographic localization of specific receptors at the level
15 of the olfactory epithelium has led to conflicting results. In some studies, electrophysiological recordings have revealed differences in olfactory responses to distinct odorant in different regions of the olfactory epithelium (34, 35). However, these experiments have been difficult to
20 interpret since the differences in response across the epithelium are often small and are not observed in all studies (36).

A second model argues that sensory neurons expressing
25 distinct odorant receptors are randomly distributed in the epithelium but that neurons responsive to a given odorant project to restricted regions within the olfactory bulb. In this instance, the discrimination of odors would be a consequence of the position of second order neurons in the
30 olfactory bulb, but would be independent of the site of origin of the afferent signals within the epithelium. Mapping of the topographic projections of olfactory neurons has been performed by extracellular recordings from different regions of the bulb (37, 38) and by 2-deoxyglucose
35 autoradiography to map regional activity after exposure to

-42-

different odorant (39). These studies suggest that spatially-localized groups of bulbar neurons preferentially respond to different odorant. The existence of specific odorant receptors, randomly distributed through the olfactory epithelium, which converge on a common target within the olfactory bulb, would raise additional questions about the recognition mechanisms used to guide these distinct axonal subsets to their central targets.

Other sensory systems also spatially segregate afferent input from primary sensory neurons. The spatial segregation of information employed, for example, by the visual and somatosensory systems, is used to define the location of the stimulus within the external environment as well as to indicate the quality of the stimulus. In contrast, olfactory processing does not extract spatial features of the odorant stimulus. Relieved of the necessity to encode information about the spatial localization of the sensory stimulus, it is possible that the olfactory system of mammals uses the spatial segregation of sensory input solely to encode the identity of the stimulus itself. The molecular identification of the genes likely to encode a large family of olfactory receptors should provide initial insights into the underlying logic of olfactory processing in the mammalian nervous system.

REFERENCES

1. Rushton, W.A.H. (1965). A Foveal pigment in the
5 deuteranope. J. Physiol. 176, 24-37.
2. Rushton, W.A.H. (1955). Foveal photopigments in normal
and colour-blind. J. Physiol. 129, 41-42.
- 10 3. Wald, G., Brown, P.K., and Smith, P.H. (1955).
Iodopsin. J. Gen. Physiol. 38, 623-681.
4. Nathans, J., Thomas, D., and Hogness, D.S. (1986).
Molecular genetics of human color vision: The genes
15 encoding blue, green and red pigments. Science 232,
193-202.
5. Lancet, D. (1986). Vertebrate olfactory reception.
Ann. Rev. Neurosci. 9, 329-355.
20
6. Reed, R.R. (1990). How does the nose know?
Minireview. Cell 60, 1-2.
7. Moulton, D.G., and Beidler, L.M. (1967). Structure and
25 function in the peripheral olfactory system. Physiol.
Rev. 47, 1-52.
8. Graziadei, P.P.C., and Monti Graziadei, G.A. (1979).
Neurogenesis and neuron regeneration in the olfactory
30 system of mammals. I. Morphological aspects of
differentiation and structural organization of the
olfactory sensory neurons. J. Neurocytol. 8, 1-18.
9. Bronshtein, A.A., and Minor, A.V. (1977). Regeneration
35 of olfactory flagella and restoration of the

-44-

electroolfactogram following application of triton X-100 to the olfactory mucosa of frogs. *Tsitologiya* 19, 33-39.

- 5 10. Rhein, L.D., and Cagan, R.H. (1980). Biochemical studies of olfaction: Isolation, characterization, and odorant binding activity of cilia from rainbow trout olfactory rosettes. *Proc. Natl. Acad. Sci. USA* 77, 4412-4416.
- 10 11. Rhein, L.D., and Cagan, R.H. (1983). Biochemical studies of olfaction: Binding specificity of odorant to a cilia preparation from rainbow trout olfactory rosettes. *J. Neurochem.* 41, 569-577.
- 15 12. Pace, U., Hanski, E., Salomon, Y., and Lancet, D. (1985). Odorant-sensitive adenylate cyclase may mediate olfactory reception. *Nature* 316, 255-258.
- 20 13. Sklar, P.B., Anholt, R.R.H., and Snyder, S.H. (1986). The odorant-sensitive adenylate cyclase of olfactory receptor cells: Differential stimulation by distinct classes of odorant. *J. Biol. Chem.* 261, 15538-15543.
- 25 14. Breer, H., Boekhoff, I., and Tareilus, E. (1990). Rapid kinetics of second messenger formation in olfactory transduction. *Nature* 345, 65-68.
- 30 15. Boekhoff, I., Tareilus, E., Strotmann, J., and Breer, H. (1990). Rapid activation of alternative second messenger pathways in olfactory cilia from rats by different odorant. *EMBO J.* 9, 2453 .
- 35 16. Jones, D.T., and Reed, R.R. (1989). G_{olf} : an olfactory neuron specific-G-protein involved in odorant signal

-45-

transduction. *Science* 244, 790-795.

- 5 17. Nakamura, T. and Gold G. (1987). A cyclic nucleotide-gated conductance in olfactory receptor cilia. *Nature* 325, 442-444.
- 10 18. Dhallan, R.S., Yau, K.-W., Schrader, K.A., and Reed, R.R. (1990). Primary structure and functional expression of a cyclic nucleotide-activated channel from olfactory neurons. *Nature* 347, 184-187.
- 15 19. O'Dowd, B.F., Felkowitz, R.J., and Caron, M.G. (1989). Structure of the adrenergic and related receptors. *Ann. Rev. Neurosci.* 12, 67-83.
- 20 20. Strader, C.D., Sigal, I.S., and Dixon, R.F. (1989). Structural basis of β -adrenergic receptor function *FASEB J.* 3, 1825-1832.
- 25 21. O'Dowd, B.F., Hnatowich, M., Caron, M.G., Lefkowitz, R.J., and Bouvier, M. (1989) Palmitoylation of the human β_2 -adrenergic receptor. Mutation of CYS³⁴¹ in the carboxyl tail leads to an uncoupled nonpalmitoylated form of the receptor. *J. Biol. Chem.* 264, 7564-7569.
- 30 22. Kobilka, B.K.I., Kobilka, T.S., Daniel, K., Regan, J.W., Caron, M.G., and Lefkowitz, R.J. (1988) Chimeric α_2 - β_2 -adrenergic receptors: delineation of domains involved in effector coupling and ligand binding specificity. *Science* 240, 1310-1316.
- 35 23. Hamm, H.E., Deretic, D., Arendt, A., Hargrove, P.A., Koenig, B., and Hofmann, K.P. (1988) Site of G protein binding to rhodopsin mapping with synthetic peptides to

-46-

the α subunit. *Science* 241, 832-835. (1988).

24. Bouvier, M.W., Hausdorff, A., De Blasi, A., O'Dowd, B.F., Kobilka, B.K., Caron, M.G., and Lefkowitz, R.J. (1988). Removal of phosphorylation sites from the β -adrenergic receptor delays the onset of agonist-promoted desensitization. *Nature* (London) 333, 370-373.
25. Amore, J.E. (1982). Odor Theory and Odor Classification, In *Fragrance Chemistry* (New York: Academic Press, Inc.), pp. 27-73.
26. Maeda, N., and Smithies, O. (1986). The evolution of multigene families: Human haptoglobin genes. *Ann. Rev. Genet.* 20, 81-108.
27. Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* 302, 575-581.
28. Hood, L., Kronenberg, M., and Hunkapiller, T. (1985). T cell antigen receptors and the immunoglobulin supergene family. *Cell* 40, 225-229.
29. Baltimore, D. (1981). Gene conversion: Some implications for immunoglobulin genes. Minireview. *Cell* 24, 592-594.
30. Egel, R. (1981). Intergenic conversion and reiterated genes. *Nature* 290, 191-192.
31. Flanagan, J.G., Lefranc, M.-P., and Rabbitts, T.H. (1984). Mechanisms of divergence and convergence of the human immunoglobulin $\alpha 1$ and $\alpha 2$ constant region gene sequences. *Cell* 36, 681-688.

-47-

32. Van der Ploeg, L.H.T. (1991). Antigenic variation in African trypanosomes: genetic recombination and transcriptional control of VSG genes. In Gene Rearrangement. (New York: Oxford University Press), pp 51-98.
33. Shepherd, G.M. (1985). Are there labeled lines in the olfactory pathway? In Taste, Olfaction and the Central Nervous System (New York: The Rockefeller University Press), pp. 307-321.
34. Mackay-Sim, A., Shaman, P., and Moulton, D.G. (1982). Topographic coding of olfactory quality: odorant-specific patterns of epithelial responsivity in the Salamander. J. Neurophysiol. 48, 584-596.
35. Thommesen, G., and Doving, K.B. (1977). Spatial distribution of the EOG in the rat. A variation with odour stimulation. Acta physiol. scand. 99, 270-280.
36. Sicard, G. (1985). Olfactory discrimination of structurally related molecules: Receptor cell responses to camphoraceous odorant. Brain Res. 326, 203-212.
37. Thommesen, G. (1978). The spatial distribution of odour induced potentials in the olfactory bulb of char and trout (Salmonidae). Acta Physiol. Scand. 102, 205-217.
38. Doving, K.B., Selset, R., and Thommesen, G. (1980). Olfactory sensitivity to bile acids in salmonid fishes. Acta Physiol. Scand. 108, 123-131.
39. Stewart, W.B., Kauer, J.S., and Shepherd, G.M. (1979).

-48-

Functional organization of the rat olfactory bulb analyzed by the 2-deoxyglucose method. J. Comp. Neurol. 185, 715-734.

- 5 40. Chirgwin, J.M., Przbyla, A.E., MacDonald, R.J., and Rutter, W.J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18, 5294-5299.
- 10 41. Maniatis, T., Fritsch, E., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).
- 15 42. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).
- 20 43. Julius et al., (1988). Molecular Characterization of a Functional cDNA Encoding a Serotonin 1c Receptor. Science 241, 558-564.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Columbia University in the City of N.Y.,
The Trustees of
- (ii) TITLE OF INVENTION: ODORANT RECEPTORS AND USES THEREOF
- (iii) NUMBER OF SEQUENCES: 36
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: COOPER & DUNHAM
 - (B) STREET: 30 Rockefeller Plaza
 - (C) CITY: New York
 - (D) STATE: New York
 - (E) COUNTRY: U.S.A.
 - (F) ZIP: 10112
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 681,880
 - (B) FILING DATE: 05-APR-1991
- (vii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: White, John P.
 - (B) REGISTRATION NUMBER: 28,678
 - (C) REFERENCE/DOCKET NUMBER: 38586
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (212) 977-9550
 - (B) TELEFAX: (212) 664-0525
 - (C) TELEX: (212) 422523 COOP UI

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 954 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: YES
- (iv) ANTI-SENSE: NO
- (v) ORIGINAL SOURCE:
 - (A) ORGANISM: rat olfactory epithelium
 - (B) STRAIN: Sprague-Dawley rat
 - (F) TISSUE TYPE: olfactory epithelium
- (vi) IMMEDIATE SOURCE:

-50-

(B) CLONE: F12

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGGAATCAG GGAACAGCAC AAGAAGATTT TCAAGTTTTT TTCTTCTTGG ATTTACAGAA	60
AACCCACAAC TTCACTTCCT CATTTTTGCA CTATTCCTGT CCATGTACCT GGTAACAGTG	120
CTTGGAACC TGCTTATCAT TATGGCCATC ATCACACAGT CTCATTTGCA TACACCCATG	180
TACTTTTTCC TTGCTAACCT ATCCTTTGTG GACATCTGTT TCACCTCCAC CACCATCCCA	240
AAGATGTTGG TAAATATATA CACCCAGAGC AAGAGCATCA CCTATGAAGA CTGTATTAGC	300
CAGATGTGTG TCTTCTTGGT TTTCGCAGAA TTGGGCAACT TTCTCCTGGC TGTGATGGCC	360
TATGACCGAT ATGTGGCTAA CTGTCACCCA CTGTGTTACA CAGTCATTGT GAACCACCGG	420
CTCTGTATCC TGCTGCTTCT GCTGTCCTGG GTTATCAGCA TTTTCCATGC CTTCATACAG	480
AGCTTAATTG TGCTACAGTT GACCTTCTGT GGAGATGTGA AAATCCCTCA CTTCTTCTGT	540
GAACCTAATC AGCTGTCCCA ACTCACCTGT TCAGACAACT TTCCAAGTCA CCTCATAATG	600
AATCTTGTA C TGTTATGTT GGCAGCCATT TCCTTCAGTG GCATCCTTTA CTCTTATTTC	660
AAGATAGTAT CCTCCATACA TTCTATCTCC ACAGTTCAGG GGAAGTACAA GGCATTTTCT	720
ACTTGTGCCT CTCACCTTTC CATTGTCTCC TTATTTTATA GTACAGGCCT CGGAGTGTAC	780
GTCAGTTCTG CTGTGGTCCA AAGCTCACAT TCTGCTGCAA GTGCTTCGGT CATGTATACT	840
GTGGTCACCC CCATGCTGAA CCCCTTCATT TATAGTCTAA GGAATAAAGA TGTGAAGAGA	900
GCTCTGGAAA GACTGTTAGA AGGAACTGT AAAGTGCATC ATTGGACTGG ATGA	954

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1002 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: F3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

-51-

ATGGACTCAA GCAACAGGAC AAGAGTTTCA GAATTTCTTC TTCTTGGATT TGTAGAAAAC	60
AAAGACCTAC AACCCCTTAT TTATGGTCTT TTTCTCTCTA TGTACCTGGT TACTGTCATT	120
GGAAACATAT CCATTATTGT GGCTATCATT TCAGATCCCT GTCTGCACAC CCCCATGTAT	180
TTCTTCCTCT CTAACCTGTC CTTTGTGGAC ATCTGTTTCA TTCAACCAC TGTCCAAAG	240
ATGTTAGTGA ACATCCAGAC CCAAACAAT GTCATCACCT ATGCAGGATG CATTACCCAG	300
ATATACTTTT TCTTGCTCTT TGTAAGATTG GACAACCTCT TGCTGACTAT CATGGCCTAT	360
GACCGTTACG TAGCCATCTG TCACCCCATG CACTACACAG TTATCATGAA CTACAAGCTC	420
TGTGGATTTC TGGTTCTGGT ATCTTGGATT GTAAGTGTTT TGCATGCCTT GTTCAAAGC	480
TTGATGATGT TGGCGCTGCC CTTCTGCACA CATCTGGAAA TCCCACACTA CTTCTGTGAA	540
CCTAATCAGG TGATTCAACT CACCTGTTCT GATGCATTTT TTAATGATCT TGTGATATAT	600
TTTACACTTG TGCTGCTGGC TACTGTTCTT CTTGCTGGCA TCTTCTATTC TTAATTCAAG	660
ATAGTGTCTT CCATATGTGC TATATCGTCA GTTCATGGGA AGTACAAAGC ATTCTCCACC	720
TGTGCATCTC ACCTTTCAGT CGTGTCTTTA TTTTACTGCA CAGGACTAGG AGTGTACCTC	780
AGTTCTGCTG CAAACAACAG CTCACAGGCA AGTGCCACAG CCTCAGTCAT GTACACTGTA	840
GTTACCCCTA TGGTGAACCC TTTTATCTAT AGTCTTAGGA ATAAAGATGT TAAGAGTGTT	900
CTGAAAAAAA CTCTTTGTGA GGAAGTTATA AGGAGTCCAC CTTCCCTACT TCATTTCTTC	960
CTAGTGTTAT GTCATCTCCC TTGTTTTATT TTTTGTATT AA	1002

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 942 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: F5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGAGCAGCA CCAACCAGTC CAGTGTACC GAGTTCCTCC TCCTGGGACT CTCCAGGCAG	60
CCCCAGCAGC AGCAGCTCCT CTTCTGCTC TTCCTCATCA TGTACCTGGC CACTGTCCTG	120

```

GGAAACCTGC TCATCATCCT GGCTATTGGC ACAGACTCCC GCCTGCACAC CCCCATGTAC      180
TTCTTCCTCA GTAACTGTC CTTTGTGGAT GTCTGCTTCT CCTCTACCAC TGTCCCTAAA      240
GTTCTGGCCA ACCATATACT TGGGAGTCAG GCCATTTCCT TCTCTGGGTG TCTCACCAG      300
CTGTATTTTC TCGCTGTGTT TGGTAACATG GACAATTTCC TGCTGGCTGT GATGTCCTAT      360
GACCGATTTG TGGCCATATG CCACCCTTTA CACTACACAA CAAAGATGAC CCGTCAGCTC      420
TGTGTCCTGC TTGTTGTGGG GTCATGGGTT GTAGCCAACA TGAATTGTCT GTTGACATA      480
CTGCTCATGG CTCGACTCTC CTTCTGTGCA GACAACATGA TCCCCCACTT CTTCTGTGAT      540
GGAACCCCC TCCTGAAACT CTCCTGCTCA GACACACATC TCAATGAGCT GATGATTCTT      600
ACAGAGGGAG CTGTGGTCAT GGTCACCCCA TTTGTCTGCA TCCTCATCTC CTACATCCAC      660
ATCACCTGTG CTGTCCTCAG AGTCTCATCC CCCAGGGGAG GATGGAAATC CTTCTCCACC      720
TGTGGCTCCC ACCTGGCTGT GGTCTGCCTC TTCTATGGCA CCGTCATCGC TGTGTATTTTC      780
AACCCATCAT CCTCTCACTT AGCTGGGAGG GACATGGCAG CTGCAGTGAT GTATGCAGTG      840
GTGACCCCAA TGCTGAACCC TTTCATCTAT AGCCTGAGGA ACAGCGACAT GAAAGCAGCT      900
TTAAGGAAAG TGCTGCCCAT GAGATTTCCA TCTAAGCAGT AA                          942

```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 936 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: F6

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

ATGGCTTGGA GTACTGGCCA GAACCTGTCC ACACCAGGAC CATTCATCTT GCTGGGCTTC      60
CCAGGGCCAA GGAGCATGCG CATTGGGCTC TTCCTGCTTT TCCTGGTCAT GTATCTGCTT      120
ACGGTAGTTG GAAACCTAGC CATCATCTCC CTGGTAGGTG CCCACAGATG CCTACAGACA      180
CCCATGTACT TCTTCCTCTG CAACCTCTCC TTCCTGGAGA TCTGGTTCAC CACAGCCTGC      240
GTACCCAAGA CCCTGGCCAC ATTTGCGCCT CGGGGTGGAG TCATTTCTT GGCTGGCTGT      300

```

-53-

```

GCCACACAGA TGTACTTTGT CTTTCTTTG GGCTGTACCG AGTACTTCCT GCTGGCTGTG   360
ATGGCTTATG ACOGCTACCT GGCCATCTGC CTGCCACTGC GCTATGGTGG CATCATGACT   420
CCTGGGCTGG CGATGCGGTT GGCCCTGGGA TCCTGGCTGT GTGGGTTTTT TGCAATCACA   480
GTTCTGCTA CCCTCATTGC CCGCCTCTCT TTCTGTGGCT CACGTGTCAT CAACCACTTC   540
TTCTGTGACA TTTCGCCCTG GATAGTGCTT TCCTGCACCG ACACGCAGGT GGTGGAACGT   600
GTGTCCTTTG GCATTGCCTT CTGTGTTATT CTGGGCTCGT GTGGTATCAC ACTAGTCTCC   660
TATGCTTACA TCATCACTAC CATCATCAAG ATTCCCTCTG CCGGGGGCCG GCACCGCGCC   720
TTCTCAACCT GCTCATCCCA TCTCACTGTG GTGCTGATTT GGTATGGCTC CACCATCTTC   780
TTGCATGTGA GGACCTCGGT AGAGAGCTCC TTGGACCTCA CCAAAGCTAT CACAGTGCTC   840
AACACCATTG TCACACCTGT GCTGAACCCT TTCATATATA CTCTGAGGAA CAAGGATGTC   900
AAGGAAGCTC TGCCAGGAC GGTGAAGGGG AAGTGA   936

```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 939 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: I14

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```

ATGACTGGAA ATAACCAAAC TTTGATCTTG GAGTTCCTCC TCCTGGGTCT GCCCATCCCA   60
TCAGAGTATC ATCTCCTGTT CTATGCCCTG TTCCTGGCCA TGTACCTCAC CATCATCCTG   120
GGAAACCTGC TAATCATTGT CCTTGTTTGA CTGGACTCTC ATCTCCACAT GCCCATGTAC   180
TTGTTTCTCA GCAACTTGTC CTTCTCTGAC CTCTGCTTTT CCTCTGTCAC AATGCCCAA   240
TTGCTTCAGA ACATGCAGAG CCAAGTACCA TCTATATCCT ATACAGGCTG CCTGACACAG   300
CTGTACTTCT TTATGGTTTT TGGAGATATG GAGAGCTTCC TTCTTGTTGGT CATGGCCTAT   360
GACCGCTATG TGGCCATTG CTTTCCTTTG CGTTACACCA CCATCATGAG CACCAAGTTC   420
TGTGCTTCAC TAGTGCTACT TCTGTGGATG CTGACGATGA CCCATGCCCT GCTGCATACC   480

```

CTACTCATTG CTAGATTGTC TTTTGTGAG AAGAATGTGA TTCTTCACTT TTTCTGTGAC 540
ATTTCTGCTC TTCTGAAGTT GTCCTGCTCA GACATTTATG TTAATGAGCT GATGATATAT 600
ATCTTGGGTG GACTCATCAT TATTATCCCA TTCCTATTAA TTGTTATGTC CTATGTTAGA 660
ATTTTCTTCT CCATTTTGAA GTTTCATCT ATTCAGGACA TCTACAAGGT ATTCTCAACC 720
TGTGGTTCCC ATCTGTCTGT GGTGACCTTG TTTTATGGGA CAATTTTGG TATCTACTTA 780
TGTCCATCAG GTAATAATTC TACTGTGAAG GAGATTGCCA TGGCTATGAT GTACACAGTG 840
GTGACTCCCA TGCTGAATCC CTTTCTCTAC AGCCTGAGGA ACAGAGACAT GAAAAGGGCC 900
CTAATAAGAG TTATCTGCAC TAAGAAAATC TCTCTGTAA 939

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 945 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: I15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

ATGACAGAAG AGAACCAAAC TGTGATCTCC CAGTTCCTTC TCCTTTTCCT GCCCATCCCC 60
TCAGAGCACC AGCACGTGTT CTACGCCCTG TTCCTGTCCA TGTACCTCAC CACTGTCCTG 120
GGGAACCTCA TCATCATCAT CCTCATTAC CTGGACTCCC ATCTCCACAC ACCCATGTAC 180
TTGTTTCTCA GCAACTTGTC CTTCTCTGAT CTCTGCTTTT CCTCTGTTAC GATGCCCAAG 240
TTGTTGCAGA ACATGCAGAG CCAAGTTCCA TCCATCCCCT TTGCAGGCTG CCTGACACAA 300
TTATACTTTT ACCTGTATTT TGCAGACCTT GAGAGCTTCC TGCTTGTGGC CATGGCCTAT 360
GACCGCTATG TGGCCATCTG CTTCCCCCTT CATTACATGA GCATCATGAG CCCCAAGCTC 420
TGTGTGAGTC TGGTGGTGCT GTCCTGGGTG CTGACCACCT TCCATGCCAT GCTGCACACC 480
CTGCTCATGG CCAGATTGTC ATTCTGTGCG GACAATATGA TCCCCCACTT TTTCTGTGAT 540
ATATCTCCTT TATTGAAACT GTCCTGCTCT GACACGCATG TTAATGAGTT GGTGATATTT 600

-55-

GTCATGGGAG GGCTTGTTAT TGTCATTCCA TTTGTGCTCA TCATTGTATC TTATGCACGA	660
GTTGTGCGCT CCATTCTTAA AGTCCCTTCT GTCCGAGGCA TCCACAAGAT CTTCTCCACC	720
TGCGGCTCCC ATCTGTCTGT GGTGTCACTG TTCTATGGGA CAATCATTGG TCTCTACTTA	780
TGTCCGTCAG CTAATAACTC TACTGTGAAG GAGACTGTCA TGGCCATGAT GTACACAGTG	840
GTGACCCCCA TGCTGAACCC CTTCATCTAC AGCCTGAGGA ACAGAGACAT GAAAGAGGCA	900
CTGATAAGAG TCCTTTGTAA AAAGAAATT ACCTTCTGTC TATGA	945

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 933 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: I3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATGAACAATC AAACCTTCAT CACCCAATTC CTTCTCCTGG GACTGCCCAT CCCTGAAGAA	60
CATCAGCACC TGTTCTATGC CTTGTTCCCTG GTCATGTACC TCACCACCAT CTTGGGAAAC	120
TTGCTAATCA TTGTACTTGT TCAACTGGAC TCCCAGCTCC ACACACCTAT GTATTTGTTT	180
CTCAGCAATT TGTCTTTCTC TGATCTATGT TTTTCCTCTG TCACAATGCC CAAGCTGCTG	240
CAGAACATGA GGAGCCAGGA CACATCCATT CCCTATGGAG GCTGCCTGGC ACAAACATAC	300
TTCTTTATGG TTTTGGAGA TATGGAGAGT TTCCTTCTTG TGGCCATGGC CTATGACCGC	360
TATGTGGCCA TCTGCTTCCC TCTGCATTAC ACCAGCATCA TGAGCCCCAA GCTCTGTACT	420
TGTCTAGTGC TGTTATTGTG GATGCTGACG ACATCCCATG CCATGATGCA CACACTGCTT	480
GCAGCAAGAT TGTCTTTTGT TGAGAACAAT GTGGTCCTCA ACTTCTTCTG TGACCTATTT	540
GTTCTCCTAA AGCTGGCCTG CTCAGACACT TATATTAATG AGTTGATGAT ATTTATCATG	600
AGTACACTCC TCATTATTAT TCCATTCTTC CTCATTGTTA TGTCTATGC AAGGATCATA	660
TCCTCTATTC TTAAGGTTC ATCTACCCAA GGCATCTGCA AGGTCTTCTC TACCTGTGGT	720

TCCCATCTGT CTGTAGTATC ACTGTTCTAT GGGACAATTA TTGGTCTCTA CTTATGTCCA	780
GCAGGTAATA ATTCCACTGT AAAAGAGATG GTCATGGCCA TGATGTACAC TGTGGTGACC	840
CCCATGCTGA ATCCCTTCAT CTACAGCCTA AGGAATAGAG ATATGAAGAG GGCCCTAATA	900
AGAGTTATCT GTAGTATGAA AATCACTCTG TAA	933

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 984 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: I7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATGGAGCGAA GGAACCACAG TGGGAGACTG AGTGAATTTG TGTGCTGGG TTTCCAGCT	60
CCTGCCCCAC TGGGAGTACT ACTATTTTTC CTTTCTCTTC TGGACTATGT GTTGGTGTTG	120
ACTGAAAACA TGCTCATCAT TATAGCAATT AGGAACCACC CAACCCTCCA CAAACCCATG	180
TATTTTTTCT TGGCTAATAT GTCATTTCTG GAGATTGGT ATGTCACTGT TACGATTCCT	240
AAGATGCTCG CTGGCTTCAT TGGTTCCAAG GAGAACCATG GACAGCTGAT CTCCTTTGAG	300
GCATGCATGA CACAACTCTA CTTTTTCCTG GGCTTGGGTT GCACAGAGTG TGTCTTCTT	360
GCTGTGATGG CCTATGACCG CTATGTGGCT ATCTGTCATC CACTCCACTA CCCCCTCATT	420
GTCAGTAGCC GGCTATGTGT GCAGATGGCA GCTGGATCCT GGGCTGGAGG TTTTGGTATC	480
TCCATGGTTA AAGTTTTCTT TATTTCTCGC CTGTCTTACT GTGGCCCCAA CACCATCAAC	540
CACTTTTTCT GTGATGTGTC TCCATTGCTC AACCTGTCAT GCACTGACAT GTCCACAGCA	600
GAGCTTACAG ACTTTGTCCT GGCCATTTTT ATTCTGCTGG GACCGCTCTC TGTCAGTGGG	660
GCATCCTACA TGGCCATCAC AGGTGCTGTG ATGCGCATCC CCTCAGCTGC TGGCCGCCAT	720
AAAGCCTTTT CAACCTGTGC CTCCACCTC ACTGTTGTGA TCATCTTCTA TGCAGCCAGT	780
ATTTTCATCT ATGCCAGGCC TAAGGCACTC TCAGCTTTTG ACACCAACAA GCTGGTCTCT	840
GTACTCTACG CTGTCATTGT ACCGTTGTTT AATCCCATCA TCTACTGCTT GCGCAACCAA	900

-57-

GATGTCAAAA GAGCGCTACG TCGCACGCTG CACCTGGCCC AGGACCAGGA GGCCAATACC 960
 AACAAAGGCA GCAAATTGG TTAG 984

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 939 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: 18

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATGAACAACA AAAGTGTGAT CACCCATTTC CTCCTCCTGG GATTGCCCCAT CCCCCCAGAG 60
 CACCAGCAAC TGTTCCTTGC CCTGTTCTTG ATCATGTACC TCACCACCTT TCTGGGAAAC 120
 CTGCTAATTG TTGTCCTTGT TCAACTGGAC TCTCATCTCC ACACACCCAT GTACTTGTTT 180
 CTCAGCAACT TGTCCTTCTC TGATCTCTGC TTTTCCTCTG TTACAATGCT GAAATTGCTG 240
 CAAAATATAC AGAGCCAAGT ACCATCTATA TCCTATGCAG GATGCCTGAC ACAGATATTC 300
 TTCTTTTTGT TGTTTGGCTA CCTTGGGAAT TTCCTTCTTG TAGCCATGGC CTATGACCGC 360
 TATGTGGCCA TCTGCTTCCC TCTGCATTAT ACCAACATCA TGAGCCATAA GCTCTGTACT 420
 TGTCTCCTGC TGGTATTTTG GATAATGACA TCATCTCATG CCATGATGCA CACCCTGCTT 480
 GCAGCAAGAT TGTCTTTTTG TGAGAACAAT GTACTCCTCA ACTTTTTCTG TGACCTGTTT 540
 GTTCTCCTAA AGTTGGCCTG CTCAGACACT TATGTTAATG AGTTGATGAT ACATATCATG 600
 GCGGTGATCA TCATTGTTAT TCCATTCTGT CTCATTGTTA TATCCTATGC CAAGATCATC 660
 TCCTCCATTC TTAAGGTTCC ATCTACTCAA AGCATTCACA AGGTCTTCTC CACTTGTTGGT 720
 TCTCATCTCT CTGTGGTGTC TCTGTTCTAC GGGACAATTA TTGGTCTCTA TTTATGTCCA 780
 TCAGGTGATA ATTTTAGTCT AAAGGGGTCT GCCATGGCTA TGATGTACAC AGTGGTAACT 840
 CCAATGCTGA ACCCGTTCAT CTACAGCCTA AGAAACAGAG ACATGAAGCA GGCCCTAATA 900
 AGAGTTACCT GTAGCAAGAA AATCTCTCTG CCATGGTAG 939

(2) INFORMATION FOR SEQ ID NO:10:

SUBSTITUTE SHEET

-58-

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 945 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: 19

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATGACTAGAA GAAACCAAAC TGCCATCTCT CAGTTCTTCC TTCTGGGCCT GCCATTCCCC	60
CCAGAGTACC AACACCTGTT CTATGCCCTG TTCCTGGCCA TGTACCTCAC CACTCTCCTG	120
GGGAACCTCA TCATCATCAT CCTCATTCTA CTGGACTCCC ATCTCCACAC ACCCATGTAC	180
TTGTTTCTCA GCAATTTATC CTTTGCOGAC CTCTGTTTTT CCTCTGTCAC AATGCCCAAG	240
TTGTTGCAGA ACATGCAGAG CCAAGTTCCA TCCATCCCCT ATGCAGGGTG CCTGGCACAG	300
ATATACTTCT TTCTGTTTTT TGGAGACCTT GGAAACTTCC TGCTTGTGGC CATGGCCTAT	360
GACCGCTATG TGGCCATCTG CTTCCCCCTT CATTACATGA GCATCATGAG CCCCAGCTC	420
TGTGTGAGTC TGGTGGTGCT GTCCTGGGTG CTGACTACCT TCCATGCCAT GCTGCACACC	480
CTGCTCATGG CCAGATTGTC ATTCTGTGAG GACAGTGTGA TCCCTCACTA TTTCTGTGAT	540
ATGTCTACTC TGCTGAAAGT GGCTTGTTC T GACACCCATG ATAATGAATT AGCAATATTT	600
ATCTTAGGGG GCCCTATAGT TGTACTACCT TTCCTTCTCA TCATTGTTTC TTATGCAAGA	660
ATTGTTTCCT CCATCTTCAA GGTCCCTTCT TCTCAAAGCA TCCATAAAGC CTTCTCCACC	720
TGTGGCTCCC ACCTGTCTGT GGTGTCACTG TTCTATGGGA CAGTCATTGG TCTCTACTTA	780
TGTCCTTCAG CTAATAACTC CACTGTGAAG GAGACTGTCA TGTCTTTGAT GTACACAATG	840
GTGACACCCA TGCTGAACCC CTTCATCTAC AGCCTAAGAA ACAGAGACAT AAAAGATGCA	900
TTAGAAAAAA TAATGTGCAA AAAGCAAATT CCCTCCTTTC TATGA	945

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 645 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

-59-

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapien

(vii) IMMEDIATE SOURCE:

(B) CLONE: H5

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..645

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATC TGT TTT GTG TCT ACC ACT GTC CCA AAG CAG CTG GTG AAC ATC CAG	48
Ile Cys Phe Val Ser Thr Thr Val Pro Lys Gln Leu Val Asn Ile Gln	
1 5 10 15	
ACA CAG AGC AGA GTC ATC ACC TAT GCA GAC TGC ATC ACC CAG ATG TGC	96
Thr Gln Ser Arg Val Ile Thr Tyr Ala Asp Cys Ile Thr Gln Met Cys	
20 25 30	
TTT TTT ATA CTC TTT GTA GTG TTG GAC AGC TTA CTC CTG ACT GTG ATG	144
Phe Phe Ile Leu Phe Val Val Leu Asp Ser Leu Leu Leu Thr Val Met	
35 40 45	
GCC TAT GAC CGG TTT GTG GCC ATC TGT CAC CCC CTG CAC TAC ACA GTC	192
Ala Tyr Asp Arg Phe Val Ala Ile Cys His Pro Leu His Tyr Thr Val	
50 55 60	
ATT ATG AGC TCC TGG CTC TGT GGA CTG CTG GTT CTG GTG TCC TGG ATC	240
Ile Met Ser Ser Trp Leu Cys Gly Leu Leu Val Leu Val Ser Trp Ile	
65 70 75 80	
GTG AGC ATC CTA TAT TCT CTG TTA CAA AGC ATA ATG GCA TTG CAG CTG	288
Val Ser Ile Leu Tyr Ser Leu Leu Gln Ser Ile Met Ala Leu Gln Leu	
85 90 95	
TCC TTC TGT ACA GAA CTG AAA ATC CCT CAA TTT TTC TGT GAA CTT AAT	336
Ser Phe Cys Thr Glu Leu Lys Ile Pro Gln Phe Phe Cys Glu Leu Asn	
100 105 110	
CAG GTC ATC CAC CTT GCC TGT TCC GAC ACT TTT ATT AAT GAC ATG ATG	384
Gln Val Ile His Leu Ala Cys Ser Asp Thr Phe Ile Asn Asp Met Met	
115 120 125	
ATG AAT TTT ACA AGT GTG CTG CTG GGT GGG GGA TGC CTC GCT GGA ATA	432
Met Asn Phe Thr Ser Val Leu Leu Gly Gly Gly Cys Leu Ala Gly Ile	
130 135 140	
TTT TAC TNN TAC TTT AAG ATA CTT TGT TGC ATA TGT TCG ATC TCA TCA	480
Phe Tyr Xaa Tyr Phe Lys Ile Leu Cys Cys Ile Cys Ser Ile Ser Ser	
145 150 155 160	
GCT CAG GGG ATG AAT AAA GCA CTT TCC ACC TGT GCA TCT CAC CTC TCA	528
Ala Gln Gly Met Asn Lys Ala Leu Ser Thr Cys Ala Ser His Leu Ser	
165 170 175	

SUBSTITUTE SHEET

-60-

GTT GTC TCC TTA TTT TAT TGT ACA GGC GTA GGT GTG TAC CTT AGT TCT	576
Val Val Ser Leu Phe Tyr Cys Thr Gly Val Gly Val Tyr Leu Ser Ser	
180 185 190	
GCT GCA ACC CAT AAC TCA CTC TCA AAT GCT GCA GCC TCG GTG ATG TAC	624
Ala Ala Thr His Asn Ser Leu Ser Asn Ala Ala Ala Ser Val Met Tyr	
195 200 205	
ACT GTG GTC ACC TCC ATG CTG	645
Thr Val Val Thr Ser Met Leu	
210 215	

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 215 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Ile Cys Phe Val Ser Thr Thr Val Pro Lys Gln Leu Val Asn Ile Gln	
1 5 10 15	
Thr Gln Ser Arg Val Ile Thr Tyr Ala Asp Cys Ile Thr Gln Met Cys	
20 25 30	
Phe Phe Ile Leu Phe Val Val Leu Asp Ser Leu Leu Leu Thr Val Met	
35 40 45	
Ala Tyr Asp Arg Phe Val Ala Ile Cys His Pro Leu His Tyr Thr Val	
50 55 60	
Ile Met Ser Ser Trp Leu Cys Gly Leu Leu Val Leu Val Ser Trp Ile	
65 70 75 80	
Val Ser Ile Leu Tyr Ser Leu Leu Gln Ser Ile Met Ala Leu Gln Leu	
85 90 95	
Ser Phe Cys Thr Glu Leu Lys Ile Pro Gln Phe Phe Cys Glu Leu Asn	
100 105 110	
Gln Val Ile His Leu Ala Cys Ser Asp Thr Phe Ile Asn Asp Met Met	
115 120 125	
Met Asn Phe Thr Ser Val Leu Leu Gly Gly Gly Cys Leu Ala Gly Ile	
130 135 140	
Phe Tyr Xaa Tyr Phe Lys Ile Leu Cys Cys Ile Cys Ser Ile Ser Ser	
145 150 155 160	
Ala Gln Gly Met Asn Lys Ala Leu Ser Thr Cys Ala Ser His Leu Ser	
165 170 175	
Val Val Ser Leu Phe Tyr Cys Thr Gly Val Gly Val Tyr Leu Ser Ser	
180 185 190	
Ala Ala Thr His Asn Ser Leu Ser Asn Ala Ala Ala Ser Val Met Tyr	
195 200 205	

SUBSTITUTE SHEET

-61-

Thr Val Val Thr Ser Met Leu
210 215

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 640 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

(B) CLONE: J1

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..640

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

C ATC TGC TTT ACT TCT GCT AGC ATC CCA AAG ATG CTA GTG AAT ATA	46
Ile Cys Phe Thr Ser Ala Ser Ile Pro Lys Met Leu Val Asn Ile	
1 5 10 15	
CAG ACG AAG AAC AAG GTG ATC ACC TAT GAA GGC TGC ATC TCC CAA GTA	94
Gln Thr Lys Asn Lys Val Ile Thr Tyr Glu Gly Cys Ile Ser Gln Val	
20 25 30	
TAC TTT TCA TAC TCT TTG GAG TTT TGG ACA ACT TTC TTC TCG ACT GTG	142
Tyr Phe Ser Tyr Ser Leu Glu Phe Trp Thr Thr Phe Phe Ser Thr Val	
35 40 45	
ATG GCC TAT GAC CGA TAT GTG GCC ATC TGT CAC CCA TCT NAC TAC ACA	190
Met Ala Tyr Asp Arg Tyr Val Ala Ile Cys His Pro Ser Xaa Tyr Thr	
50 55 60	
GGT CAT CAT GAA CCN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN	238
Gly His His Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
65 70 75	
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN	286
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
80 85 90 95	
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN	334
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
100 105 110	
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN	382
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	

SUBSTITUTE SHEET

-62-

115	120	125	
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NTT			430
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa			
130	135	140	
TAT TCT TAC TCT AAG ATA GTT TCC TCC ATA CGA GAA ATC TCA TCA TCA			478
Tyr Ser Tyr Ser Lys Ile Val Ser Ser Ile Arg Glu Ile Ser Ser Ser			
145	150	155	
CAG GGA AAG TAC AAG NNA TTC TCC ACC TGT GCA TCC CAC CTC TCA GTT			526
Gln Gly Lys Tyr Lys Xaa Phe Ser Thr Cys Ala Ser His Leu Ser Val			
160	165	170	175
GTT TCA TTA TTC TAT TCT ACA CTT TTG GGT GTG TAC CTT AGT TCT TCT			574
Val Ser Leu Phe Tyr Ser Thr Leu Leu Gly Val Tyr Leu Ser Ser Ser			
180	185	190	
TTT ACC CAA AAC TCA CAC TCA ACT GCA CGG GCA TCT GTT ATG TAC AGT			622
Phe Thr Gln Asn Ser His Ser Thr Ala Arg Ala Ser Val Met Tyr Ser			
195	200	205	
GTG GTC ACC CCC ATG TTG			640
Val Val Thr Pro Met Leu			
210			

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 213 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Ile Cys Phe Thr Ser Ala Ser Ile Pro Lys Met Leu Val Asn Ile Gln			
1	5	10	15
Thr Lys Asn Lys Val Ile Thr Tyr Glu Gly Cys Ile Ser Gln Val Tyr			
20	25	30	
Phe Ser Tyr Ser Leu Glu Phe Trp Thr Thr Phe Phe Ser Thr Val Met			
35	40	45	
Ala Tyr Asp Arg Tyr Val Ala Ile Cys His Pro Ser Xaa Tyr Thr Gly			
50	55	60	
His His Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa			
65	70	75	80
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa			
85	90	95	
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa			
100	105	110	
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa			
115	120	125	

-63-

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Tyr
 130 135 140
 Ser Tyr Ser Lys Ile Val Ser Ser Ile Arg Glu Ile Ser Ser Ser Gln
 145 150 155 160
 Gly Lys Tyr Lys Xaa Phe Ser Thr Cys Ala Ser His Leu Ser Val Val
 165 170 175
 Ser Leu Phe Tyr Ser Thr Leu Leu Gly Val Tyr Leu Ser Ser Ser Phe
 180 185 190
 Thr Gln Asn Ser His Ser Thr Ala Arg Ala Ser Val Met Tyr Ser Val
 195 200 205
 Val Thr Pro Met Leu
 210

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 636 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (iii) HYPOTHETICAL: YES
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: rat olfactory epithelium
 - (B) STRAIN: sprague-Dawley rat
 - (F) TISSUE TYPE: olfactory epithelium
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: J2
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..636
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

ACC TCC ACC ACC ATC CCA AAG ATG CTG GTA AAT ATA CAC ACC CAG AGC	48
Thr Ser Thr Thr Ile Pro Lys Met Leu Val Asn Ile His Thr Gln Ser	
1 5 10 15	
AAT ACT ATC ACC TAT GAA GAC TGT ATT TCC CAG ATG TTT GTA CTC TTG	96
Asn Thr Ile Thr Tyr Glu Asp Cys Ile Ser Gln Met Phe Val Leu Leu	
20 25 30	
GTT TTT GGA GAA CTG GAC AAC TTT CTC CTG GCT GTG ATG GCC TAT GAT	144
Val Phe Gly Glu Leu Asp Asn Phe Leu Leu Ala Val Met Ala Tyr Asp	
35 40 45	
CGA TAT GTG GCT ATC TGT CAC CCA CTG TAT TAC ACA GTC ATT GTG AAC	192
Arg Tyr Val Ala Ile Cys His Pro Leu Tyr Tyr Thr Val Ile Val Asn	
50 55 60	

SUBSTITUTE SHEET

-64-

CAC CGA CTC TGT ATC CTG CTG CTT CTG CTG TCC TGG GTT GTC AGC ATT	240
His Arg Leu Cys Ile Leu Leu Leu Leu Leu Ser Trp Val Val Ser Ile	
65 70 75 80	
TTA CAT GCC TTC TTA CAG AGC TTA ATT GTA CTA CAG TTG ACC TTC TGT	288
Leu His Ala Phe Leu Gln Ser Leu Ile Val Leu Gln Leu Thr Phe Cys	
85 90 95	
GGA GAT GTG AAA ATC CCT CAC TTC TTC TGT GAG CTC AAT CAG CTG TCC	336
Gly Asp Val Lys Ile Pro His Phe Phe Cys Glu Leu Asn Gln Leu Ser	
100 105 110	
CAA CTC ACA TGT TCA GAC AAC TTT CCA AGT CAC CTC ACA ATG CAT CTT	384
Gln Leu Thr Cys Ser Asp Asn Phe Pro Ser His Leu Thr Met His Leu	
115 120 125	
GTA CCT GTT ATA TTT GCA GCT ATT TCC CTC AGT GGT ATC CTT TAC TCT	432
Val Pro Val Ile Phe Ala Ala Ile Ser Leu Ser Gly Ile Leu Tyr Ser	
130 135 140	
TAT TTC AAG ATA GTG TCT TCC ATA CGT TCT ATG TCC TCA GTT CAA GGG	480
Tyr Phe Lys Ile Val Ser Ser Ile Arg Ser Met Ser Ser Val Gln Gly	
145 150 155 160	
AAG TAC AAG GCA TTT TCT ACA TGT GCC TCT CAC CTT TCC ATT GTC TCC	528
Lys Tyr Lys Ala Phe Ser Thr Cys Ala Ser His Leu Ser Ile Val Ser	
165 170 175	
TTA TTT TAT AGT ACA GGC CTC GGG GTG TAC GTC AGT TCT GCT GTG ATC	576
Leu Phe Tyr Ser Thr Gly Leu Gly Val Tyr Val Ser Ser Ala Val Ile	
180 185 190	
CGA AGC TCA CAC TCC TCT GCA AGT GCT TCG GTC ATG TAT ACT GTG GTC	624
Arg Ser Ser His Ser Ser Ala Ser Ala Ser Val Met Tyr Thr Val Val	
195 200 205	
ACC CCC ATG TTG	636
Thr Pro Met Leu	
210	

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 212 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Thr Ser Thr Thr Ile Pro Lys Met Leu Val Asn Ile His Thr Gln Ser	
1 5 10 15	
Asn Thr Ile Thr Tyr Glu Asp Cys Ile Ser Gln Met Phe Val Leu Leu	
20 25 30	
Val Phe Gly Glu Leu Asp Asn Phe Leu Leu Ala Val Met Ala Tyr Asp	
35 40 45	
Arg Tyr Val Ala Ile Cys His Pro Leu Tyr Tyr Thr Val Ile Val Asn	

-65-

50		55		60
His Arg Leu Cys Ile Leu Leu Leu Leu Ser Trp Val Val Ser Ile				
65		70		75 80
Leu His Ala Phe Leu Gln Ser Leu Ile Val Leu Gln Leu Thr Phe Cys				
	85		90	95
Gly Asp Val Lys Ile Pro His Phe Phe Cys Glu Leu Asn Gln Leu Ser				
	100		105	110
Gln Leu Thr Cys Ser Asp Asn Phe Pro Ser His Leu Thr Met His Leu				
	115		120	125
Val Pro Val Ile Phe Ala Ala Ile Ser Leu Ser Gly Ile Leu Tyr Ser				
	130		135	140
Tyr Phe Lys Ile Val Ser Ser Ile Arg Ser Met Ser Ser Val Gln Gly				
145		150		155 160
Lys Tyr Lys Ala Phe Ser Thr Cys Ala Ser His Leu Ser Ile Val Ser				
	165		170	175
Leu Phe Tyr Ser Thr Gly Leu Gly Val Tyr Val Ser Ser Ala Val Ile				
	180		185	190
Arg Ser Ser His Ser Ser Ala Ser Ala Ser Val Met Tyr Thr Val Val				
	195		200	205
Thr Pro Met Leu				
210				

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 646 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: erpague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: J4

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..646

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

C ATA GGC TAT TCA TCT TCT GTC ACA CCC AAT ATG CTT GTC AAC TTC

46

-66-

Ile	Gly	Tyr	Ser	Ser	Ser	Val	Thr	Pro	Asn	Met	Leu	Val	Asn	Phe		
1				5					10					15		
CTT	ATA	AAG	CAA	AAT	ACC	ATC	TCA	TAC	CTT	GGA	TGT	TCT	ATA	CAG	TTT	94
Leu	Ile	Lys	Gln	Asn	Thr	Ile	Ser	Tyr	Leu	Gly	Cys	Ser	Ile	Gln	Phe	
				20					25					30		
GGC	TCA	GCT	GCT	TTG	TTT	GGA	GGT	CTT	GAA	TGC	TTC	CTT	CTG	GCT	GCC	142
Gly	Ser	Ala	Ala	Leu	Phe	Gly	Gly	Leu	Glu	Cys	Phe	Leu	Leu	Ala	Ala	
				35				40					45			
ATG	GCG	TAT	GAT	CGT	TTT	GTA	GCA	ATC	TGC	AAC	CCA	CTG	CTT	TAT	TCA	190
Met	Ala	Tyr	Asp	Arg	Phe	Val	Ala	Ile	Cys	Asn	Pro	Leu	Leu	Tyr	Ser	
				50				55					60			
ACG	AAA	ATG	TCC	ACA	CAA	GTC	TGT	GTC	CAG	TTG	GTT	GTG	GGA	TCT	TAT	238
Thr	Lys	Met	Ser	Thr	Gln	Val	Cys	Val	Gln	Leu	Val	Val	Gly	Ser	Tyr	
				65				70					75			
ATA	GCG	GGA	TTT	CTT	AAT	GCC	TCC	TCT	TTT	ACC	CTT	TCC	TTT	TTT	TCC	286
Ile	Gly	Gly	Phe	Leu	Asn	Ala	Ser	Ser	Phe	Thr	Leu	Ser	Phe	Phe	Ser	
						85					90				95	
TTG	TCC	TTC	TGT	GGA	CCA	AAT	AGA	ATC	AAT	CAC	TTT	TAC	TGT	GAT	TTT	334
Leu	Ser	Phe	Cys	Gly	Pro	Asn	Arg	Ile	Asn	His	Phe	Tyr	Cys	Asp	Phe	
				100					105					110		
GCT	CCG	TTA	GTA	GAA	CTT	TCT	TGC	TCT	GAT	GTC	AGT	GTT	CCT	GAT	GCT	382
Ala	Pro	Leu	Val	Glu	Leu	Ser	Cys	Ser	Asp	Val	Ser	Val	Pro	Asp	Ala	
				115				120					125			
GTT	ACC	TCA	TTT	TCT	GCT	GCC	TCA	GTT	ACT	ATG	CTC	ACA	GTG	TTT	ATC	430
Val	Thr	Ser	Phe	Ser	Ala	Ala	Ser	Val	Thr	Met	Leu	Thr	Val	Phe	Ile	
				130				135					140			
ATA	GCC	ATC	TCC	TAT	ACC	TAT	ATC	CTC	ATC	ACC	ATC	CTG	AAG	ATG	CGT	478
Ile	Ala	Ile	Ser	Tyr	Thr	Tyr	Ile	Leu	Ile	Thr	Ile	Leu	Lys	Met	Arg	
				145				150					155			
TCC	ACT	GAG	GGT	CGA	CAG	AAA	GCA	TTC	TCT	ACC	TGC	ACT	TCC	CAC	CTC	526
Ser	Thr	Glu	Gly	Arg	Gln	Lys	Ala	Phe	Ser	Thr	Cys	Thr	Ser	His	Leu	
				160				165					170		175	
ACT	GCA	GTC	ACT	CTG	TGC	TAT	GGA	ACC	ATC	ACA	TTC	ATC	TAT	GTG	ATG	574
Thr	Ala	Val	Thr	Leu	Cys	Tyr	Gly	Thr	Ile	Thr	Phe	Ile	Tyr	Val	Met	
				180					185					190		
CCC	AAG	TCC	AGC	TAC	TCC	ACA	GAC	CAG	AAC	AAG	GTG	GTG	TCT	GTG	TTT	622
Pro	Lys	Ser	Ser	Tyr	Ser	Thr	Asp	Gln	Asn	Lys	Val	Val	Ser	Val	Phe	
				195				200					205			
TAT	ATG	GTG	GTG	ATC	CCC	ATG	TTG									646
Tyr	Met	Val	Val	Ile	Pro	Met	Leu									
				210			215									

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 215 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

-67-

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```

Ile Gly Tyr Ser Ser Ser Val Thr Pro Asn Met Leu Val Asn Phe Leu
 1              5              10              15
Ile Lys Gln Asn Thr Ile Ser Tyr Leu Gly Cys Ser Ile Gln Phe Gly
              20              25              30
Ser Ala Ala Leu Phe Gly Gly Leu Glu Cys Phe Leu Leu Ala Ala Met
              35              40              45
Ala Tyr Asp Arg Phe Val Ala Ile Cys Asn Pro Leu Leu Tyr Ser Thr
              50              55              60
Lys Met Ser Thr Gln Val Cys Val Gln Leu Val Val Gly Ser Tyr Ile
              65              70              75              80
Gly Gly Phe Leu Asn Ala Ser Ser Phe Thr Leu Ser Phe Phe Ser Leu
              85              90              95
Ser Phe Cys Gly Pro Asn Arg Ile Asn His Phe Tyr Cys Asp Phe Ala
              100              105              110
Pro Leu Val Glu Leu Ser Cys Ser Asp Val Ser Val Pro Asp Ala Val
              115              120              125
Thr Ser Phe Ser Ala Ala Ser Val Thr Met Leu Thr Val Phe Ile Ile
              130              135              140
Ala Ile Ser Tyr Thr Tyr Ile Leu Ile Thr Ile Leu Lys Met Arg Ser
              145              150              155              160
Thr Glu Gly Arg Gln Lys Ala Phe Ser Thr Cys Thr Ser His Leu Thr
              165              170              175
Ala Val Thr Leu Cys Tyr Gly Thr Ile Thr Phe Ile Tyr Val Met Pro
              180              185              190
Lys Ser Ser Tyr Ser Thr Asp Gln Asn Lys Val Val Ser Val Phe Tyr
              195              200              205
Met Val Val Ile Pro Met Leu
              210              215

```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 481 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium

SUBSTITUTE SHEET

-68-

(B) STRAIN: Sprague-Dawley rat
(F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:
(B) CLONE: J7

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..481

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

C ATC TGC AAG CCC CTG CAC TAC ACC ACC ATC ATG AAT AAC CGA GTG	46
Ile Cys Lys Pro Leu His Tyr Thr Thr Ile Met Asn Asn Arg Val	
1 5 10 15	
TGC ACA GTT CTA GTC CTC TCC TGT TGG TTT GCT GGC CTG TTG ATC ATC	94
Cys Thr Val Leu Val Leu Ser Cys Trp Phe Ala Gly Leu Leu Ile Ile	
20 25 30	
CTC CCA CCT CTT GGT CAT GGC CTC CAG CTG GAG TTC TGT GAC TCC AAT	142
Leu Pro Pro Leu Gly His Gly Leu Gln Leu Glu Phe Cys Asp Ser Asn	
35 40 45	
GTG ATT GAT CAT TTT GGC TGT GAT GCC TCT CCA ATT CTG CAG ATA ACC	190
Val Ile Asp His Phe Gly Cys Asp Ala Ser Pro Ile Leu Gln Ile Thr	
50 55 60	
TGC TCA GAC ACG GTA TTT ATA GAG AAA ATT GTC TTG GCT TTT GCC ATA	238
Cys Ser Asp Thr Val Phe Ile Glu Lys Ile Val Leu Ala Phe Ala Ile	
65 70 75	
CTG ACA CTC ATC ATT ACT CTG GTA TGT GTT GTT CTC TCC TAC ACA TAC	286
Leu Thr Leu Ile Ile Thr Leu Val Cys Val Val Leu Ser Tyr Thr Tyr	
80 85 90 95	
ATC ATC AAG ACC ATT TTA AAG TTT CCT TCT GCT CAA CAA AGA AAA AAG	334
Ile Ile Lys Thr Ile Leu Lys Phe Pro Ser Ala Gln Gln Arg Lys Lys	
100 105 110	
GCC TTT TCT ACA TGT TCT TCC CAC ATG ATT GTG GTT TCC ATC ACC TAT	382
Ala Phe Ser Thr Cys Ser Ser His Met Ile Val Val Ser Ile Thr Tyr	
115 120 125	
GCG AGC TGT ATT TTC ATC TAC ATC AAA CCT TCA GCG AAG GAA GGG GTA	430
Gly Ser Cys Ile Phe Ile Tyr Ile Lys Pro Ser Ala Lys Glu Gly Val	
130 135 140	
GCC ATC AAT AAG GTT GTA TCT GTG CTC ACA ACA TCA GTC GCC CCT TTG	478
Ala Ile Asn Lys Val Val Ser Val Leu Thr Thr Ser Val Ala Pro Leu	
145 150 155	
CTC	481
Leu	
160	

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 160 amino acids

-69-

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```

Ile Cys Lys Pro Leu His Tyr Thr Thr Ile Met Asn Asn Arg Val Cys
 1           5           10           15
Thr Val Leu Val Leu Ser Cys Trp Phe Ala Gly Leu Leu Ile Ile Leu
 20           25           30
Pro Pro Leu Gly His Gly Leu Gln Leu Glu Phe Cys Asp Ser Asn Val
 35           40           45
Ile Asp His Phe Gly Cys Asp Ala Ser Pro Ile Leu Gln Ile Thr Cys
 50           55           60
Ser Asp Thr Val Phe Ile Glu Lys Ile Val Leu Ala Phe Ala Ile Leu
 65           70           75           80
Thr Leu Ile Ile Thr Leu Val Cys Val Val Leu Ser Tyr Thr Tyr Ile
 85           90           95
Ile Lys Thr Ile Leu Lys Phe Pro Ser Ala Gln Gln Arg Lys Lys Ala
 100          105          110
Phe Ser Thr Cys Ser Ser His Met Ile Val Val Ser Ile Thr Tyr Gly
 115          120          125
Ser Cys Ile Phe Ile Tyr Ile Lys Pro Ser Ala Lys Glu Gly Val Ala
 130          135          140
Ile Asn Lys Val Val Ser Val Leu Thr Thr Ser Val Ala Pro Leu Leu
 145          150          155          160

```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 481 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: rat olfactory epithelium
 (B) STRAIN: Sprague-Dawley rat
 (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:
 (B) CLONE: J8

(ix) FEATURE:
 (A) NAME/KEY: CDS

SUBSTITUTE SHEET

-70-

(B) LOCATION: 2..481

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:21:

C	ATC	TGC	CAC	CCG	CTC	CAC	TAC	TCT	CTT	CTC	ATG	AGT	CCT	GAC	AAC	46
Ile	Cys	His	Pro	Leu	His	Tyr	Ser	Leu	Leu	Met	Ser	Pro	Asp	Asn		
1				5				10					15			
TGT	GCT	GCT	CTG	GTA	ACA	GTC	TCC	TGG	GTG	ACA	GGG	GTG	GGC	ACG	GGC	94
Cys	Ala	Ala	Leu	Val	Thr	Val	Ser	Trp	Val	Thr	Gly	Val	Gly	Thr	Gly	
			20					25					30			
TTC	CTG	CCT	TCC	CTC	CTG	ATT	TCT	AAG	TTG	GAC	TTC	TGT	GGG	CCC	AAC	142
Phe	Leu	Pro	Ser	Leu	Leu	Ile	Ser	Lys	Leu	Asp	Phe	Cys	Gly	Pro	Asn	
			35					40					45			
CGC	ATC	AAC	CAT	TTC	TTC	TGT	GAC	CTC	CCT	CCA	TTA	ATC	CAG	CTG	TCC	190
Arg	Ile	Asn	His	Phe	Phe	Cys	Asp	Leu	Pro	Pro	Leu	Ile	Gln	Leu	Ser	
		50					55					60				
TGC	TCC	AGC	GTC	TTT	GTG	ACA	GAA	ATG	GCC	ATC	TTT	GTC	CTG	TCC	ATC	238
Cys	Ser	Ser	Val	Phe	Val	Thr	Glu	Met	Ala	Ile	Phe	Val	Leu	Ser	Ile	
	65					70					75					
GCT	GTG	CTC	TGC	ATC	TGT	TTC	CTC	CTA	ACC	CNN	NNN	TCC	TAC	ATT	TTC	286
Ala	Val	Leu	Cys	Ile	Cys	Phe	Leu	Leu	Thr	Xaa	Xaa	Ser	Tyr	Ile	Phe	
	80				85				90					95		
ATA	GTG	TCC	TCC	ATT	CTG	AGA	ATC	CCT	TCC	ACT	ACC	GGC	AGG	ATG	AAG	334
Ile	Val	Ser	Ser	Ile	Leu	Arg	Ile	Pro	Ser	Thr	Thr	Gly	Arg	Met	Lys	
				100				105						110		
ACA	TTT	TCT	ACA	TGT	GGC	TCC	CAC	CTG	GCC	GTG	GTC	ACC	ATC	TAC	TAT	382
Thr	Phe	Ser	Thr	Cys	Gly	Ser	His	Leu	Ala	Val	Val	Thr	Ile	Tyr	Tyr	
			115					120					125			
GGG	ACC	ATG	ATC	TCC	ATG	TAT	GTC	GCC	CCA	AAT	GCG	CAT	CTG	TCC	CCG	430
Gly	Thr	Met	Ile	Ser	Met	Tyr	Val	Gly	Pro	Asn	Ala	His	Leu	Ser	Pro	
		130					135					140				
GAG	CTC	AAC	AAG	GTC	ATT	TCT	GTC	TTC	TAC	ACT	GTG	ATC	ACC	CCA	CTA	478
Glu	Leu	Asn	Lys	Val	Ile	Ser	Val	Phe	Tyr	Thr	Val	Ile	Thr	Pro	Leu	
	145					150					155					
CTG																481
Leu																
160																

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 160 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Ile Cys His Pro Leu His Tyr Ser Leu Leu Met Ser Pro Asp Asn Cys

-71-

1	5	10	15												
Ala	Ala	Leu	Val	Thr	Val	Ser	Trp	Val	Thr	Gly	Val	Gly	Thr	Gly	Phe
		20						25					30		
Leu	Pro	Ser	Leu	Leu	Ile	Ser	Lys	Leu	Asp	Phe	Cys	Gly	Pro	Asn	Arg
		35					40					45			
Ile	Asn	His	Phe	Phe	Cys	Asp	Leu	Pro	Pro	Leu	Ile	Gln	Leu	Ser	Cys
	50					55					60				
Ser	Ser	Val	Phe	Val	Thr	Glu	Met	Ala	Ile	Phe	Val	Leu	Ser	Ile	Ala
	65				70					75					80
Val	Leu	Cys	Ile	Cys	Phe	Leu	Leu	Thr	Xaa	Xaa	Ser	Tyr	Ile	Phe	Ile
			85					90						95	
Val	Ser	Ser	Ile	Leu	Arg	Ile	Pro	Ser	Thr	Thr	Gly	Arg	Met	Lys	Thr
			100				105						110		
Phe	Ser	Thr	Cys	Gly	Ser	His	Leu	Ala	Val	Val	Thr	Ile	Tyr	Tyr	Gly
		115					120					125			
Thr	Met	Ile	Ser	Met	Tyr	Val	Gly	Pro	Asn	Ala	His	Leu	Ser	Pro	Glu
	130					135					140				
Leu	Asn	Lys	Val	Ile	Ser	Val	Phe	Tyr	Thr	Val	Ile	Thr	Pro	Leu	Leu
	145				150					155					160

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 646 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: J11

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..646

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

N	GTC	TGC	TTC	TCC	TCC	ACC	ACT	GTC	CCC	AAG	GTA	CTG	GCT	AAC	CAC
	Val	Cys	Phe	Ser	Ser	Thr	Thr	Val	Pro	Lys	Val	Leu	Ala	Asn	His
	1				5					10				15	

46

-72-

ATA CTC AGT AGT CAG GCC ATT TCC TTC TCT GGG TGT CTA ACT CAG CTG Ile Leu Ser Ser Gln Ala Ile Ser Phe Ser Gly Cys Leu Thr Gln Leu 20 25 30	94
TAT TTT CTC TGT GTG TCT GTG AAT ATG GAC AAT TTC CTG CTG GCT GTG Tyr Phe Leu Cys Val Ser Val Asn Met Asp Asn Phe Leu Leu Ala Val 35 40 45	142
ATG GCC TAT GAC AGA TTT GTG GCC ATA TGC CAC CCT TTG TAC TAC ACA Met Ala Tyr Asp Arg Phe Val Ala Ile Cys His Pro Leu Tyr Tyr Thr 50 55 60	190
ACA AAG ATG ACC CAC CAG CTC TGT GTC TTG CTG GTG TCT GGA TCA NNN Thr Lys Met Thr His Gln Leu Cys Val Leu Leu Val Ser Gly Ser Xaa 65 70 75	238
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 80 85 90 95	286
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 100 105 110	334
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 115 120 125	382
NNN NNN NNN NNN NNN NNN NNT GTG ATC ATG GTC ACC CCA TTT GTC TGC Xaa Xaa Xaa Xaa Xaa Xaa Xaa Val Ile Met Val Thr Pro Phe Val Cys 130 135 140	430
ATC CTC ATC TCT TAC ATC TAC ATC ACC AAT GCA GTC CTC AGA GTC TCA Ile Leu Ile Ser Tyr Ile Tyr Ile Thr Asn Ala Val Leu Arg Val Ser 145 150 155	478
TCC TTT AGG GGA GGA TGG AAA GCC TTC TCC ACC TGT GGC TCA CAC CTG Ser Phe Arg Gly Gly Trp Lys Ala Phe Ser Thr Cys Gly Ser His Leu 160 165 170 175	526
GCT GTG GTC TGC CTC TTC TAT GGC ACC ATC ATT GCT GTG TAT TTC AAT Ala Val Val Cys Leu Phe Tyr Gly Thr Ile Ile Ala Val Tyr Phe Asn 180 185 190	574
CCT GTA TCT TCC CAT TCA TCT GAG AAG GAC ACT GCA GCA ACT GTG CTA Pro Val Ser Ser His Ser Ser Glu Lys Asp Thr Ala Ala Thr Val Leu 195 200 205	622
TAC ACA GTG GTG ACT CCC ATG TTG Tyr Thr Val Val Thr Pro Met Leu 210 215	646

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 215 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

-73-

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Val Cys Phe Ser Ser Thr Thr Val Pro Lys Val Leu Ala Asn His Ile
 1 5 10 15
 Leu Ser Ser Gln Ala Ile Ser Phe Ser Gly Cys Leu Thr Gln Leu Tyr
 20 25 30
 Phe Leu Cys Val Ser Val Asn Met Asp Asn Phe Leu Leu Ala Val Met
 35 40 45
 Ala Tyr Asp Arg Phe Val Ala Ile Cys His Pro Leu Tyr Tyr Thr Thr
 50 55 60
 Lys Met Thr His Gln Leu Cys Val Leu Leu Val Ser Gly Ser Xaa Xaa
 65 70 75 80
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 85 90 95
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 100 105 110
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 115 120 125
 Xaa Xaa Xaa Xaa Xaa Xaa Val Ile Met Val Thr Pro Phe Val Cys Ile
 130 135 140
 Leu Ile Ser Tyr Ile Tyr Ile Thr Asn Ala Val Leu Arg Val Ser Ser
 145 150 155 160
 Phe Arg Gly Gly Trp Lys Ala Phe Ser Thr Cys Gly Ser His Leu Ala
 165 170 175
 Val Val Cys Leu Phe Tyr Gly Thr Ile Ile Ala Val Tyr Phe Asn Pro
 180 185 190
 Val Ser Ser His Ser Ser Glu Lys Asp Thr Ala Ala Thr Val Leu Tyr
 195 200 205
 Thr Val Val Thr Pro Met Leu
 210 215

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 646 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

-74-

(vii) IMMEDIATE SOURCE:
(B) CLONE: J14

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..646

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

T GTC TGC TTC TCC TCC ACC ACT GTC CCC AAG GTA CTG GCT AAC CAC	46
Val Cys Phe Ser Ser Thr Thr Val Pro Lys Val Leu Ala Asn His	
1 5 10 15	
ATA CTC AGT AGT CAG GCC ATT TCC TTC TCT GGG TGT CTA ACT CAG CTG	94
Ile Leu Ser Ser Gln Ala Ile Ser Phe Ser Gly Cys Leu Thr Gln Leu	
20 25 30	
TAT TTT CTC TGT GTG TCT GTG AAT ATG GAC AAT TTC CTG CTG GCT CTG	142
Tyr Phe Leu Cys Val Ser Val Asn Met Asp Asn Phe Leu Leu Ala Val	
35 40 45	
ATG GCC TAT GAC AGA TTT GTG GCC ATA TGC CAC CCT TTG TAC TAC ACA	190
Met Ala Tyr Asp Arg Phe Val Ala Ile Cys His Pro Leu Tyr Tyr Thr	
50 55 60	
ACA CCG ATG ACC CAC CAG CTC TGT GTC TTG CTG GTG TCT GGA TCA NNN	238
Thr Pro Met Thr His Gln Leu Cys Val Leu Leu Val Ser Gly Ser Xaa	
65 70 75	
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN	286
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
80 85 90 95	
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN	334
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
100 105 110	
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN	382
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
115 120 125	
NNN NNN NNN NNN NNN NNN NNT GTG ATC ATG GTC ACC CCA TTT GTC TGC	430
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Val Ile Met Val Thr Pro Phe Val Cys	
130 135 140	
ATC CTC ATC TCT TAC ATC TAC ATC ACC AAT GCA GTC CTC AGA GTC TCA	478
Ile Leu Ile Ser Tyr Ile Tyr Ile Thr Asn Ala Val Leu Arg Val Ser	
145 150 155	
TCC TTT AGG GGA GGA TGG AAA GCC TTC TCC ACC TGT GGC TCA CAC CTG	526
Ser Phe Arg Gly Gly Trp Lys Ala Phe Ser Thr Cys Gly Ser His Leu	
160 165 170 175	
GCT GTG GTC TGC CTC TTC TAT GGC ACC ATC ATT GCT GTG TAT TTC AAT	574
Ala Val Val Cys Leu Phe Tyr Gly Thr Ile Ile Ala Val Tyr Phe Asn	
180 185 190	
CCT GTA TCT TCC CAT TCA TCT GAG AAG GAC ACT GCA GCA ACT GTG CTA	622
Pro Val Ser Ser His Ser Ser Glu Lys Asp Thr Ala Ala Thr Val Leu	
195 200 205	

-75-

TAC ACA GTG GTG ACT CCC ATG TTG
 Tyr Thr Val Val Thr Pro Met Leu
 210 215

646

(2) INFORMATION FOR SEQ ID NO:26:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 215 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Val Cys Phe Ser Ser Thr Thr Val Pro Lys Val Leu Ala Asn His Ile
 1 5 10 15
 Leu Ser Ser Gln Ala Ile Ser Phe Ser Gly Cys Leu Thr Gln Leu Tyr
 20 25 30
 Phe Leu Cys Val Ser Val Asn Met Asp Asn Phe Leu Leu Ala Val Met
 35 40 45
 Ala Tyr Asp Arg Phe Val Ala Ile Cys His Pro Leu Tyr Tyr Thr Thr
 50 55 60
 Pro Met Thr His Gln Leu Cys Val Leu Leu Val Ser Gly Ser Xaa Xaa
 65 70 75 80
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 85 90 95
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 100 105 110
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 115 120 125
 Xaa Xaa Xaa Xaa Xaa Xaa Val Ile Met Val Thr Pro Phe Val Cys Ile
 130 135 140
 Leu Ile Ser Tyr Ile Tyr Ile Thr Asn Ala Val Leu Arg Val Ser Ser
 145 150 155 160
 Phe Arg Gly Gly Trp Lys Ala Phe Ser Thr Cys Gly Ser His Leu Ala
 165 170 175
 Val Val Cys Leu Phe Tyr Gly Thr Ile Ile Ala Val Tyr Phe Asn Pro
 180 185 190
 Val Ser Ser His Ser Ser Glu Lys Asp Thr Ala Ala Thr Val Leu Tyr
 195 200 205
 Thr Val Val Thr Pro Met Leu
 210 215

(2) INFORMATION FOR SEQ ID NO:27:

(1) SEQUENCE CHARACTERISTICS:

-76-

(A) LENGTH: 481 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: rat olfactory epithelium
 (B) STRAIN: Sprague-Dawley rat
 (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

(B) CLONE: J15

(ix) FEATURE:

(A) NAME/KEY: CDS
 (B) LOCATION: 2..481

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

T ATC TGC AAC CCT CTG CGC TAC CCA GTG CTC ATG AGC GGC CGG GTG	46
Ile Cys Asn Pro Leu Arg Tyr Pro Val Leu Met Ser Gly Arg Val	
1 5 10 15	
TGC CTG CTC ATG GTC GTG GCC TCC TGG TTG GGA GGA TCC CTC AAC GCC	94
Cys Leu Leu Met Val Val Ala Ser Trp Leu Gly Gly Ser Leu Asn Ala	
20 25 30	
TCC ATT CAG ACT TCT CTG ACC CTT CAG TTC CCC TAC TGT GGA TCA CGG	142
Ser Ile Gln Thr Ser Leu Thr Leu Gln Phe Pro Tyr Cys Gly Ser Arg	
35 40 45	
AAG ATC TCC CAC TTC TTC TGT GAG GTG CCC TCG CTG CTG ANN NTG GCC	190
Lys Ile Ser His Phe Phe Cys Glu Val Pro Ser Leu Leu Xaa Xaa Ala	
50 55 60	
TGT GCA GAC ACT GAA GCC TAT GAG CAG GTA CTA TTT GTG ACA GGC GTG	238
Cys Ala Asp Thr Glu Ala Tyr Glu Gln Val Leu Phe Val Thr Gly Val	
65 70 75	
GTG GTC CTC CTG GTG CCC ATT ACA TTC ATT ACT GCC TCT TAT GCC CTC	286
Val Val Leu Leu Val Pro Ile Thr Phe Ile Thr Ala Ser Tyr Ala Leu	
80 85 90 95	
ATC CTG GCT GCT GTG CTC CGA ATG CAC TCT GCG GAG GGG AGT CAG AAG	334
Ile Leu Ala Ala Val Leu Arg Met His Ser Ala Glu Gly Ser Gln Lys	
100 105 110	
GCC CTA GCC ACA TGC TCC TCT CAC CTG ACA GTC GTC AAT CTC TTC TAT	382
Ala Leu Ala Thr Cys Ser Ser His Leu Thr Val Val Asn Leu Phe Tyr	
115 120 125	
GGG CCC CTT GTC TAC ACC TAC ATG TTA CCT GCT TCC TAT CAC TCA CCA	430
Gly Pro Leu Val Tyr Thr Tyr Met Leu Pro Ala Ser Tyr His Ser Pro	
130 135 140	

-77-

GGC CAA GAC GAC ATA GTA TCC GTC TTT TAC ACC GTT CTC ACA CCC ATG 478
 Gly Gln Asp Asp Ile Val Ser Val Phe Tyr Thr Val Leu Thr Pro Met
 145 150 155

CTT 481
 Leu
 160

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 160 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Ile Cys Asn Pro Leu Arg Tyr Pro Val Leu Met Ser Gly Arg Val Cys
 1 5 10 15
 Leu Leu Met Val Val Ala Ser Trp Leu Gly Gly Ser Leu Asn Ala Ser
 20 25 30
 Ile Gln Thr Ser Leu Thr Leu Gln Phe Pro Tyr Cys Gly Ser Arg Lys
 35 40 45
 Ile Ser His Phe Phe Cys Glu Val Pro Ser Leu Leu Xaa Xaa Ala Cys
 50 55 60
 Ala Asp Thr Glu Ala Tyr Glu Gln Val Leu Phe Val Thr Gly Val Val
 65 70 75 80
 Val Leu Leu Val Pro Ile Thr Phe Ile Thr Ala Ser Tyr Ala Leu Ile
 85 90 95
 Leu Ala Ala Val Leu Arg Met His Ser Ala Glu Gly Ser Gln Lys Ala
 100 105 110
 Leu Ala Thr Cys Ser Ser His Leu Thr Val Val Asn Leu Phe Tyr Gly
 115 120 125
 Pro Leu Val Tyr Thr Tyr Met Leu Pro Ala Ser Tyr His Ser Pro Gly
 130 135 140
 Gln Asp Asp Ile Val Ser Val Phe Tyr Thr Val Leu Thr Pro Met Leu
 145 150 155 160

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 481 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

-78-

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: rat olfactory epithelium

(B) STRAIN: Sprague-Dawley rat

(F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

(B) CLONE: J16

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 2..481

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

C ATC TGT AGG CCT CTT CAC TAT CCT ACC CTC ATG ACC CAG ACA CTG	46
Ile Cys Arg Pro Leu His Tyr Pro Thr Leu Met Thr Gln Thr Leu	
1 5 10 15	
TGT GCC AAG ATT GCC ACT GGT TGC TGG TTG GGA GGC TTG GCT GGG CCA	94
Cys Ala Lys Ile Ala Thr Gly Cys Trp Leu Gly Gly Leu Ala Gly Pro	
20 25 30	
GTG GTA GAA ATT TCC TTG GTG TCT CGT CTC CTT TTT TGT GGC CCC AAT	142
Val Val Glu Ile Ser Leu Val Ser Arg Leu Leu Phe Cys Gly Pro Asn	
35 40 45	
CAC ATT CAA CAC ATC TTT TGT GAT TTC CCA CCT GTG CTG AGC TTG GCT	190
His Ile Gln His Ile Phe Cys Asp Phe Pro Pro Val Leu Ser Leu Ala	
50 55 60	
TGT ACT GAT ACA TCA GTG AAT GTC CTG GTA GAT TTT ATT ATA AAC CTC	238
Cys Thr Asp Thr Ser Val Asn Val Leu Val Asp Phe Ile Ile Asn Leu	
65 70 75	
TGC AAG ATC CTG GCC ACC TTC CTG CTG ATC CTG AGC TCC TAC TTG CAG	286
Cys Lys Ile Leu Ala Thr Phe Leu Leu Ile Leu Ser Ser Tyr Leu Gln	
80 85 90 95	
ATA ATC CGC ACA GTG CTC AAG ATT CCT TCA GCT GCA GGC AAG AAG AAA	334
Ile Ile Arg Thr Val Leu Lys Ile Pro Ser Ala Ala Gly Lys Lys Lys	
100 105 110	
GCA TTC TCG ACT TGT GCC TCC CAT CTC ACT GTG GTT CTC ATC TTC TAT	382
Ala Phe Ser Thr Cys Ala Ser His Leu Thr Val Val Leu Ile Phe Tyr	
115 120 125	
GGG AGC ATC CTT TTC ATG TAT GTG CGG CTG AAG AAG ACT TAC TCC CTT	430
Gly Ser Ile Leu Phe Met Tyr Val Arg Leu Lys Lys Thr Tyr Ser Leu	
130 135 140	
GAC TAC GAC AGA GCC TTG GCA GTA GTC TAC TCC GTG GTT ACC CCT TTC	478
Asp Tyr Asp Arg Ala Leu Ala Val Val Tyr Ser Val Val Thr Pro Phe	
145 150 155	
CTG	481
Leu	
160	

(2) INFORMATION FOR SEQ ID NO:30:

-79-

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 160 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```

Ile Cys Arg Pro Leu His Tyr Pro Thr Leu Met Thr Gln Thr Leu Cys
 1           5           10           15
Ala Lys Ile Ala Thr Gly Cys Trp Leu Gly Gly Leu Ala Gly Pro Val
          20           25           30
Val Glu Ile Ser Leu Val Ser Arg Leu Leu Phe Cys Gly Pro Asn His
          35           40           45
Ile Gln His Ile Phe Cys Asp Phe Pro Pro Val Leu Ser Leu Ala Cys
          50           55           60
Thr Asp Thr Ser Val Asn Val Leu Val Asp Phe Ile Ile Asn Leu Cys
          65           70           75           80
Lys Ile Leu Ala Thr Phe Leu Leu Ile Leu Ser Ser Tyr Leu Gln Ile
          85           90           95
Ile Arg Thr Val Leu Lys Ile Pro Ser Ala Ala Gly Lys Lys Lys Ala
          100          105          110
Phe Ser Thr Cys Ala Ser His Leu Thr Val Val Leu Ile Phe Tyr Gly
          115          120          125
Ser Ile Leu Phe Met Tyr Val Arg Leu Lys Lys Thr Tyr Ser Leu Asp
          130          135          140
Tyr Asp Arg Ala Leu Ala Val Val Tyr Ser Val Val Thr Pro Phe Leu
          145          150          155          160

```

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 481 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: J17

-80-

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 2..481

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

A	ATC	TGC	AAC	CCA	CTG	CTT	TAT	TCC	ACC	AAA	ATG	TCC	ACA	CAA	GTC	46
	Ile	Cys	Asn	Pro	Leu	Leu	Tyr	Ser	Thr	Lys	Met	Ser	Thr	Gln	Val	
	1				5					10					15	
	TGT	ATC	CAG	TTG	GTT	GCA	GGA	TCT	TAT	ATA	GGG	GGT	TTT	CTT	AAT	ACT
	Cys	Ile	Gln	Leu	Val	Ala	Gly	Ser	Tyr	Ile	Gly	Gly	Phe	Leu	Asn	Thr
				20						25					30	
	TGC	CTC	ATC	ATG	TTT	TAC	TTT	TTC	TCT	TTT	CTC	TTC	TGT	GGG	CCA	AAT
	Cys	Leu	Ile	Met	Phe	Tyr	Phe	Phe	Ser	Phe	Leu	Phe	Cys	Gly	Pro	Asn
				35					40					45		
	ATA	GTT	GAT	CAT	TTT	TTC	TGT	GAT	TTT	GCT	CCT	TTN	NTG	GAA	CTT	TCG
	Ile	Val	Asp	His	Phe	Phe	Cys	Asp	Phe	Ala	Pro	Xaa	Xaa	Glu	Leu	Ser
				50				55					60			
	TGC	TCT	GAT	GTG	AGT	GTC	TCT	GTA	GTT	GTT	ATG	TCA	TTT	TCT	GCT	GGC
	Cys	Ser	Asp	Val	Ser	Val	Ser	Val	Val	Val	Met	Ser	Phe	Ser	Ala	Gly
		65					70					75				
	TCA	GTT	ACT	ATG	ATC	ACA	GTG	TTT	ATC	ATA	GCC	ATC	TCC	TAT	TCT	TAC
	Ser	Val	Thr	Met	Ile	Thr	Val	Phe	Ile	Ile	Ala	Ile	Ser	Tyr	Ser	Tyr
		80				85					90					95
	ATC	CTC	ATC	ACC	ATC	CTG	AAG	ATG	TCC	TCA	ACT	GAG	GGC	CGT	CAC	AAG
	Ile	Leu	Ile	Thr	Ile	Leu	Lys	Met	Ser	Ser	Thr	Glu	Gly	Arg	His	Lys
					100					105					110	
	GCT	TTC	TCC	ACA	TGT	ACC	TCC	CAC	CTC	ACT	GCA	GTC	ACT	CTC	TAC	TAT
	Ala	Phe	Ser	Thr	Cys	Thr	Ser	His	Leu	Thr	Ala	Val	Thr	Leu	Tyr	Tyr
				115					120					125		
	GGC	ACC	ATT	ACC	TTC	ATT	TAT	GTG	ATG	CCC	AAG	TCC	ACA	TAC	TCT	ACA
	Gly	Thr	Ile	Thr	Phe	Ile	Tyr	Val	Met	Pro	Lys	Ser	Thr	Tyr	Ser	Thr
			130					135					140			
	GAC	CAG	AAC	AAG	GTG	GTG	TCT	GTG	TTT	TAC	ATG	GTG	GTG	ATC	CCA	ATG
	Asp	Gln	Asn	Lys	Val	Val	Ser	Val	Phe	Tyr	Met	Val	Val	Ile	Pro	Met
		145					150				155					
	TTG															481
	Leu															
	160															

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 160 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

-81-

```

Ile Cys Asn Pro Leu Leu Tyr Ser Thr Lys Met Ser Thr Gln Val Cys
 1           5           10           15
Ile Gln Leu Val Ala Gly Ser Tyr Ile Gly Gly Phe Leu Asn Thr Cys
           20           25           30
Leu Ile Met Phe Tyr Phe Phe Ser Phe Leu Phe Cys Gly Pro Asn Ile
           35           40           45
Val Asp His Phe Phe Cys Asp Phe Ala Pro Xaa Xaa Glu Leu Ser Cys
           50           55           60
Ser Asp Val Ser Val Ser Val Val Val Met Ser Phe Ser Ala Gly Ser
           65           70           75           80
Val Thr Met Ile Thr Val Phe Ile Ile Ala Ile Ser Tyr Ser Tyr Ile
           85           90           95
Leu Ile Thr Ile Leu Lys Met Ser Ser Thr Glu Gly Arg His Lys Ala
           100          105          110
Phe Ser Thr Cys Thr Ser His Leu Thr Ala Val Thr Leu Tyr Tyr Gly
           115          120          125
Thr Ile Thr Phe Ile Tyr Val Met Pro Lys Ser Thr Tyr Ser Thr Asp
           130          135          140
Gln Asn Lys Val Val Ser Val Phe Tyr Met Val Val Ile Pro Met Leu
           145          150          155          160

```

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 479 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: J19

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..479

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```

T ATC TGC CAC CCT CTG AAG TAC ACA GTT ATC ATG AAT CAC TAT TTT
Ile Cys His Pro Leu Lys Tyr Thr Val Ile Met Asn His Tyr Phe

```

46

-82-

1	5	10	15	
TGT GTG ATG CTG CTG CTC TTC TCT GTG TTC GTT AGC ATT GCA CAT GCG				94
Cys Val Met Leu Leu Leu Phe Ser Val Phe Val Ser Ile Ala His Ala	20	25	30	
TTG TTC CAC ATT TTA ATG GTG TTG ATA CTG ACT TTC AGC ACA AAA ACT				142
Leu Phe His Ile Leu Met Val Leu Ile Leu Thr Phe Ser Thr Lys Thr	35	40	45	
GAA ATC CCT CAC TTT TTC TGT GAG CTG GCT CAT ATC ATC AAA CTT ACC				190
Glu Ile Pro His Phe Phe Cys Glu Leu Ala His Ile Ile Lys Leu Thr	50	55	60	
TGT TCC GAT AAT TTT ATC AAC TAT CTG CTG ATA TAC ACA GAG TCT GTC				238
Cys Ser Asp Asn Phe Ile Asn Tyr Leu Leu Ile Tyr Thr Glu Ser Val	65	70	75	
TTA TTT TTT GGT GTT CAT ATT GTA GGG ATC ATT TTG TCT TAT ATT TAC				286
Leu Phe Phe Gly Val His Ile Val Gly Ile Ile Leu Ser Tyr Ile Tyr	80	85	90	95
ACT GTA TCC TCA GTT TTA AGA ATG TCA TTA TTG GGA GGA ATG TAT AAA				334
Thr Val Ser Ser Val Leu Arg Met Ser Leu Leu Gly Gly Met Tyr Lys	100	105	110	
GCC TTT TCA ACA TGT GGA TCT CAT TTG TCG GTT GTC TCT GTT TTA TGG				382
Ala Phe Ser Thr Cys Gly Ser His Leu Ser Val Val Ser Val Leu Trp	115	120	125	
CAC AGG TTT TGG GGT ACA CAT AAG CTC TCC ACT TAC TGA CTC TCC AAG				430
His Arg Phe Trp Gly Thr His Lys Leu Ser Thr Tyr * Leu Ser Lys	130	135	140	
GAA GAC TGT AGT GGC TTC AGT GAT GTA CAC TGT GGT TAC TCA GAT GCT G				479
Glu Asp Cys Ser Gly Phe Ser Asp Val His Cys Gly Tyr Ser Asp Ala	145	150	155	

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 159 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Ile Cys His Pro Leu Lys Tyr Thr Val Ile Met Asn His Tyr Phe Cys
 1 5 10 15
 Val Met Leu Leu Leu Phe Ser Val Phe Val Ser Ile Ala His Ala Leu
 20 25 30
 Phe His Ile Leu Met Val Leu Ile Leu Thr Phe Ser Thr Lys Thr Glu
 35 40 45
 Ile Pro His Phe Phe Cys Glu Leu Ala His Ile Ile Lys Leu Thr Cys
 50 55 60

-83-

Ser Asp Asn Phe Ile Asn Tyr Leu Leu Ile Tyr Thr Glu Ser val Leu
65 70 75 80

Phe Phe Gly Val His Ile Val Gly Ile Ile Leu Ser Tyr Ile Tyr Thr
85 90 95

Val Ser Ser Val Leu Arg Met Ser Leu Leu Gly Gly Met Tyr Lys Ala
100 105 110

Phe Ser Thr Cys Gly Ser His Leu Ser Val Val Ser Val Leu Trp His
115 120 125

Arg Phe Trp Gly Thr His Lys Leu Ser Thr Tyr * Leu Ser Lys Glu
130 135 140

Asp Cys Ser Gly Phe Ser Asp Val His Cys Gly Tyr Ser Asp Ala
145 150 155

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 481 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: J20

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..481

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

A ATC TGC TAC CCA CTG AGG TAC CTT CTC ATC ATG AGC TGG GTC GTC Ile Cys Tyr Pro Leu Arg Tyr Leu Leu Ile Met Ser Trp Val Val 1 5 10 15	46
TGC ACA GCA CTG TCC GTG GCA ATC TGG GTC ATA GGC TTT TGT GCC TCC Cys Thr Ala Leu Ser Val Ala Ile Trp Val Ile Gly Phe Cys Ala Ser 20 25 30	94
GTT ATA CCT CTC TGC TTC ACG ATC CTC CCA CTC TGT GGT CCT TAC GTC Val Ile Pro Leu Cys Phe Thr Ile Leu Pro Leu Cys Gly Pro Tyr Val 35 40 45	142
GTT GAT TAT CTT TTC TGC GAG CTG CCC ATC CTT CTG CAC CTG TTC TGC Val Asp Tyr Leu Phe Cys Glu Leu Pro Ile Leu Leu His Leu Phe Cys 50 55 60	190
ACA GAT ACA TCT CTG CTG GAG NNN NNN NNN NNN NNN NNN NNN NNN Thr Asp Thr Ser Leu Leu Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 65 70 75	238

SUBSTITUTE SHEET

-84-

NNN NNN NNN NNN NNN	CCC TTC CTC CTG ATT GTT CTC TCC TAC CTT CGC	286
Xaa Xaa Xaa Xaa Xaa	Pro Phe Leu Leu Ile Val Leu Ser Tyr Leu Arg	
80	85 90 95	
ATC CTG GTG GCT GTG ATA AGA ATA GAC TCA GCT GAG GGC AGA AAA AAG	334	
Ile Leu Val Ala Val Ile Arg Ile Asp Ser Ala Glu Gly Arg Lys Lys		
100 105 110		
GCC TTT TCA ACT TGT GCT TCA CAC TTG GCT GTG GTG ACC ATC TAC TAT	382	
Ala Phe Ser Thr Cys Ala Ser His Leu Ala Val Val Thr Ile Tyr Tyr		
115 120 125		
GGA ACA GGG CTG ATC AGG TAC TTG AGG CCC AAG TCC CTT TAT TCC GCT	430	
Gly Thr Gly Leu Ile Arg Tyr Leu Arg Pro Lys Ser Leu Tyr Ser Ala		
130 135 140		
GAG GGA GAC AGA CTG ATC TCT GTG TTC TAT GCA GTC ATT GGC CCT GCA	478	
Glu Gly Asp Arg Leu Ile Ser Val Phe Tyr Ala Val Ile Gly Pro Ala		
145 150 155		
CTG	481	
Leu		
160		

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 160 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Ile Cys Tyr Pro Leu Arg Tyr Leu Leu Ile Met Ser Trp Val Val Cys	
1 5 10 15	
Thr Ala Leu Ser Val Ala Ile Trp Val Ile Gly Phe Cys Ala Ser Val	
20 25 30	
Ile Pro Leu Cys Phe Thr Ile Leu Pro Leu Cys Gly Pro Tyr Val Val	
35 40 45	
Asp Tyr Leu Phe Cys Glu Leu Pro Ile Leu Leu His Leu Phe Cys Thr	
50 55 60	
Asp Thr Ser Leu Leu Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
65 70 75 80	
Xaa Xaa Xaa Xaa Pro Phe Leu Leu Ile Val Leu Ser Tyr Leu Arg Ile	
85 90 95	
Leu Val Ala Val Ile Arg Ile Asp Ser Ala Glu Gly Arg Lys Lys Ala	
100 105 110	
Phe Ser Thr Cys Ala Ser His Leu Ala Val Val Thr Ile Tyr Tyr Gly	
115 120 125	
Thr Gly Leu Ile Arg Tyr Leu Arg Pro Lys Ser Leu Tyr Ser Ala Glu	
130 135 140	
Gly Asp Arg Leu Ile Ser Val Phe Tyr Ala Val Ile Gly Pro Ala Leu	
145 150 155 160	

SUBSTITUTE SHEET

What is claimed is:

1. An isolated nucleic acid molecule encoding an odorant receptor.
2. An isolated DNA of claim 1.
3. An isolated cDNA of claim 2.
4. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 9.
5. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 10.
6. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 11.
7. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 12.
8. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 13.
9. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 14.
10. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 15.
11. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 16.
12. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 17.

13. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 18.
14. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 19.
15. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 20.
16. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 21.
17. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 22.
18. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 23.
19. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 24.
20. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 25.
21. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 26.
22. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 27.
23. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 28.
24. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 29.

25. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 30.
- 5 26. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 31.
27. An isolated cDNA of claim 3 encoding an insect odorant receptor.
- 10 28. An isolated cDNA of claim 3 encoding a vertebrate odorant receptor.
29. An isolated cDNA of claim 3 encoding a fish odorant receptor.
- 15 30. An isolated cDNA of claim 3 encoding a mammalian odorant receptor.
- 20 31. An isolated cDNA of claim 30 wherein the mammalian odorant receptor is a human odorant receptor.
32. An isolated cDNA of claim 30 wherein the mammalian odorant receptor is a rat, dog or mouse odorant receptor.
- 25 33. An expression vector comprising the cDNA of claim 3 and the sequence elements necessary for replication and expression in a suitable host.
- 30 34. An expression vector comprising the cDNA of any of claims 4-19 and the sequence elements necessary for replication and expression in a suitable host.
35. A purified protein encoding an odorant receptor.

36. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 9.
- 5 37. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 10.
38. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 11.
- 10 39. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 12.
40. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 13.
- 15 41. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 14.
42. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 15.
- 20 43. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 16.
- 25 44. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 17.
45. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 18.
- 30 46. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 19.
- 35 47. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 20.

-89-

48. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 21.
- 5 49. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 22.
50. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 23.
- 10 51. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 24.
52. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 25.
- 15 53. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 26.
54. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 27.
- 20 55. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 28.
- 25 56. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 29.
57. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 31.
- 30 58. A purified protein of claim 35 encoding an insect odorant receptor.
- 35 59. A purified protein of claim 35 encoding a vertebrate odorant receptor.

-90-

60. A purified protein of claim 35 encoding a fish odorant receptor.
- 5 61. A purified protein of claim 35 encoding a mammalian odorant receptor.
62. A purified protein of claim 61 wherein the mammalian odorant receptor is a human odorant receptor.
- 10 63. A purified protein of claim 61 wherein the mammalian odorant receptor is a rat, dog or mouse odorant receptor.
- 15 64. A purified protein of claim 35 which has 7 transmembrane regions and whose third cytoplasmic loop from the N-terminus is approximately 17 amino acid long.
- 20 65. A method of transforming cells which comprises transfecting a host cell with a suitable expression vector of claim 33.
- 25 66. A method of transforming cells which comprises transfecting a host cell with a suitable expression vector of claim 34.
67. Cells transformed by the method of claim 65.
- 30 68. Transformed cells of claim 67 wherein the cells are olfactory cells.
69. Transformed cells of claim 67 wherein the cells are non-olfactory cells.

35

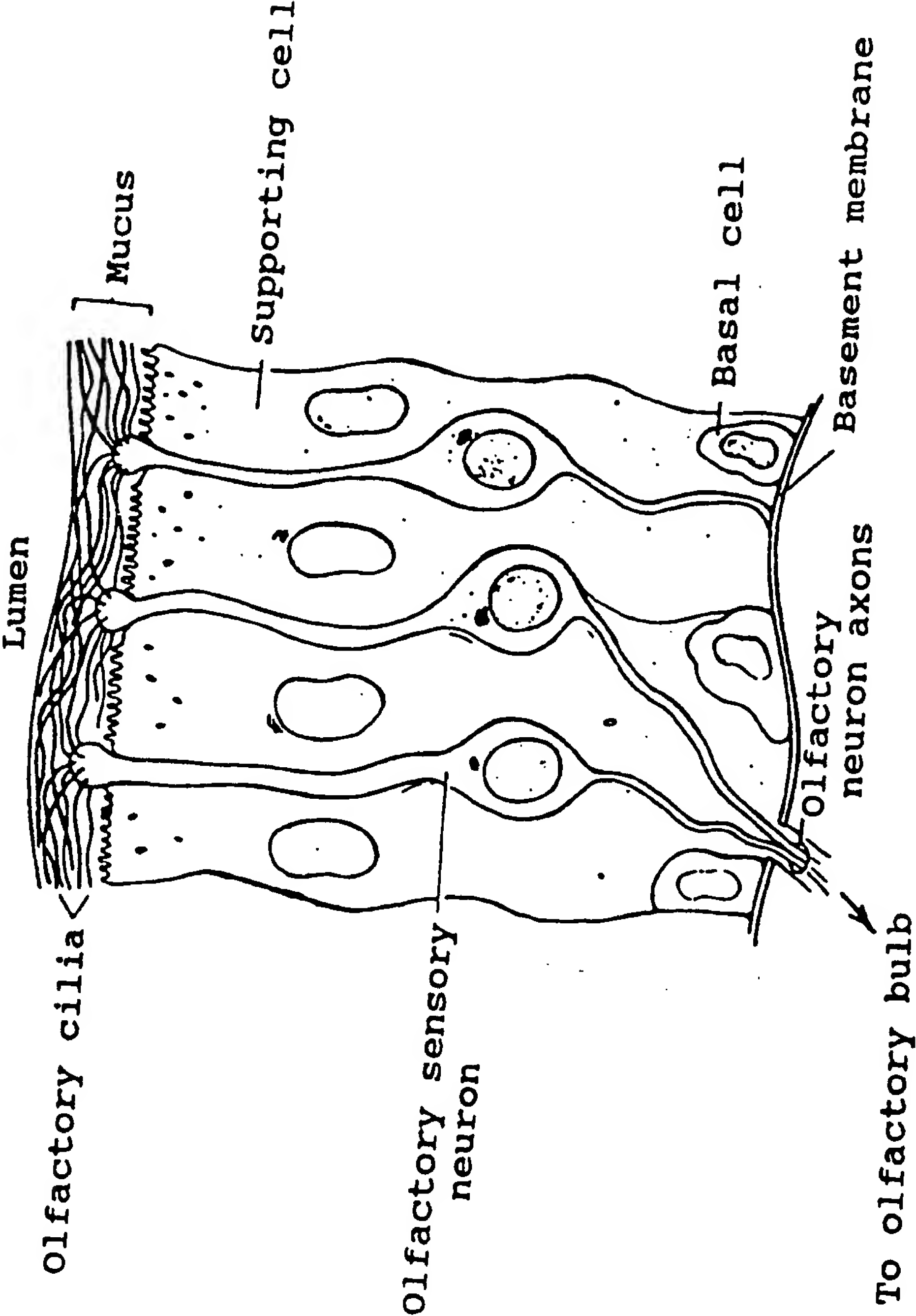
70. A method of identifying a desired odorant ligand comprising contacting transformed non-olfactory cells of claim 69, expressing a known odorant receptor with a series of odorant ligands and determining which ligands bind to the receptors present on the non-olfactory cells.
71. A method of identifying a desired odorant receptor comprising contacting a series of transformed non-olfactory cells of claim 69 with a known odorant ligand and determining which odorant receptor binds with the odorant ligand.
72. A method of detecting an odor which comprises:
- a) identifying a odorant receptor which binds the desired odorant ligand by the method of claim 71 and;
 - b) imbedding the receptor in a membrane such that when the odorant ligand binds with the receptor identified in a) above, a detectable signal is produced.
73. A method of claim 72 wherein the desired odorant is a pheromone.
74. A method of claim 72 wherein the desired odorant ligand is the vapors emanating from cocaine, marijuana, heroin, hashish, or angel dust.
75. A method of claim 72 wherein the desired odorant ligand is the vapors emanating from gasoline, natural gas or alcohol.

76. A method of claim 72 wherein the desired odorant ligand is the vapors emanating from decayed human flesh.
- 5 77. A method of claim 72 wherein the desired odorant ligand is the vapors emanating from explosives, plastic explosives, firearms, or gun powder.
- 10 78. A method of claim 72 wherein the desired odorant ligand is toxic fumes, noxious fumes or dangerous fumes.
79. A method of claim 72 wherein the membrane is a cell membrane.
- 15 80. A method of claim 72 wherein the membrane is an olfactory cell membrane.
81. A method of claim 72 wherein the membrane is a synthetic membrane.
- 20 82. A method of claim 72 wherein the detectable signal is a color change, phosphorescence, or radioactivity.
- 25 83. A method of quantifying the amount of an odorant ligand present in a sample which comprises the method of claim 72 wherein the detectable signal is quantified.
- 30 84. A method of developing fragrances which comprises identifying a desired odorant receptor by the method of claim 71 then contacting non-olfactory cells, which have been transfected with an expression vector containing the cDNA of the desired odorant receptor such that the receptor is expressed upon the surface of the non-olfactory cell, with a series of compounds to determine which ones bind with the receptor.
- 35

- 5 85. A method of identifying an odorant fingerprint which comprises contacting a series of cells, which have been transformed such that each express a known odorant receptor, with a desired sample and determining the type and quantity of the odorant ligands present in the sample.
- 10 86. A method of identifying odorant ligands which inhibit the activity of a desired odorant receptor which comprises contacting the desired odorant receptor with a series of compounds and determining which compounds inhibit the odorant ligand - odorant receptor interaction.
- 15 87. A method of identifying appetite suppressant compounds which comprises identifying odorant ligands by the method of claim 86 wherein the desired odorant receptor is that which is associated with the perception of food.
- 20 88. A method of controlling appetite in a subject which comprises contacting the olfactory epithelium of the subject with the odorant ligands identified by the method of claim 87.
- 25 89. A nasal spray, to control appetite comprising the compounds identified by the method of claim 87 in a suitable carrier.
- 30 90. A method of trapping odors which comprises contacting a membrane which contains multiples of the desired odorant receptor, with a sample such that the desired odorant ligand is absorbed by the binding of the odorant ligand to the odorant receptor.
- 35

91. An odor trap employing the method of claim 90.
- 5 92. A method of controlling pest populations which comprises identifying odorant ligands by the method of claim 70 which are alarm odorant ligands and spraying the desired area with the identified odorant ligands.
- 10 93. A method of controlling a pest population which comprises identifying odorant ligands by the method of claim 70 which interfere with the interaction between the odorant ligands and the odorant receptors which are associated with fertility.
- 15 94. A method of claim 92 or 93 wherein the pest population is a population of insects.
95. A method of claim 92 or 93 wherein the pest population is a population of rodents.
- 20 96. A method of claim 95 wherein the population of rodents is a population of mice or rats.
- 25 97. A method of promoting fertility which comprises employing the method of claim 70 to identify odorant ligands which interact with the odorant receptors associated with fertility and administering the identified odorant ligands to a subject.
- 30 98. A method of inhibiting fertility which comprises employing the method of claim 70 to identifying odorant ligands which inhibit the interaction between the odorant ligands and the odorant receptors associated with fertility and administering the identified odorant ligands to a subject.

Figure 1A



2/99

Figure 1B

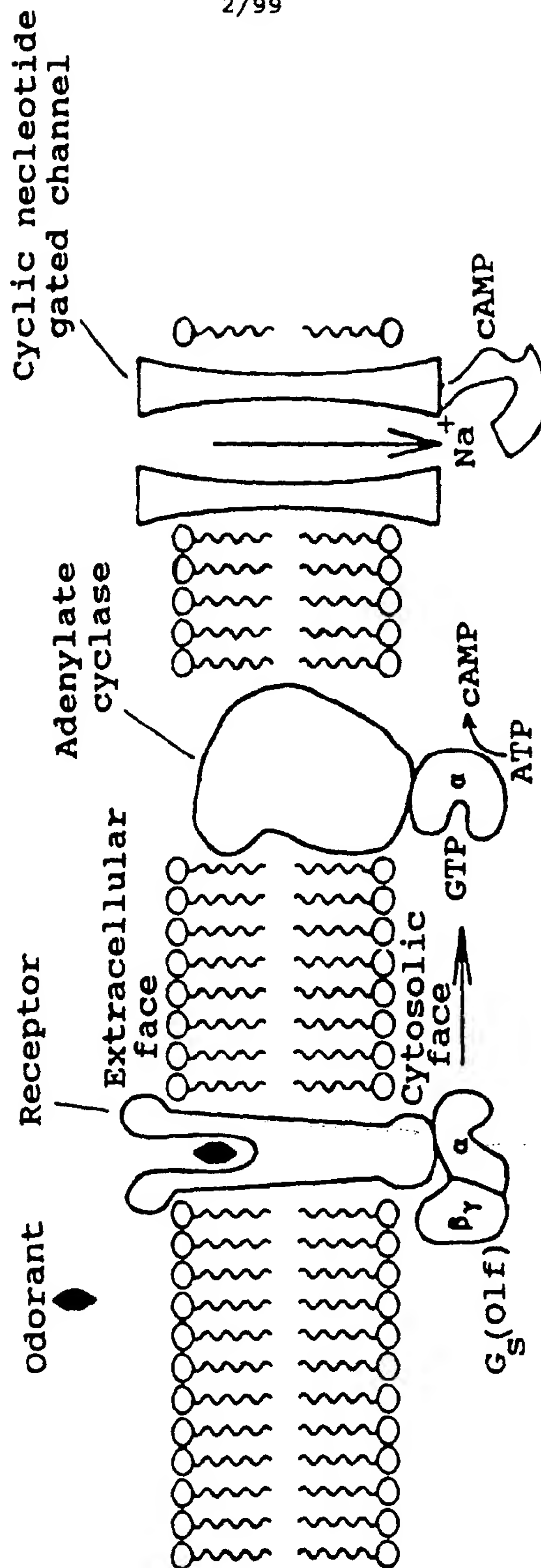


Figure 2A

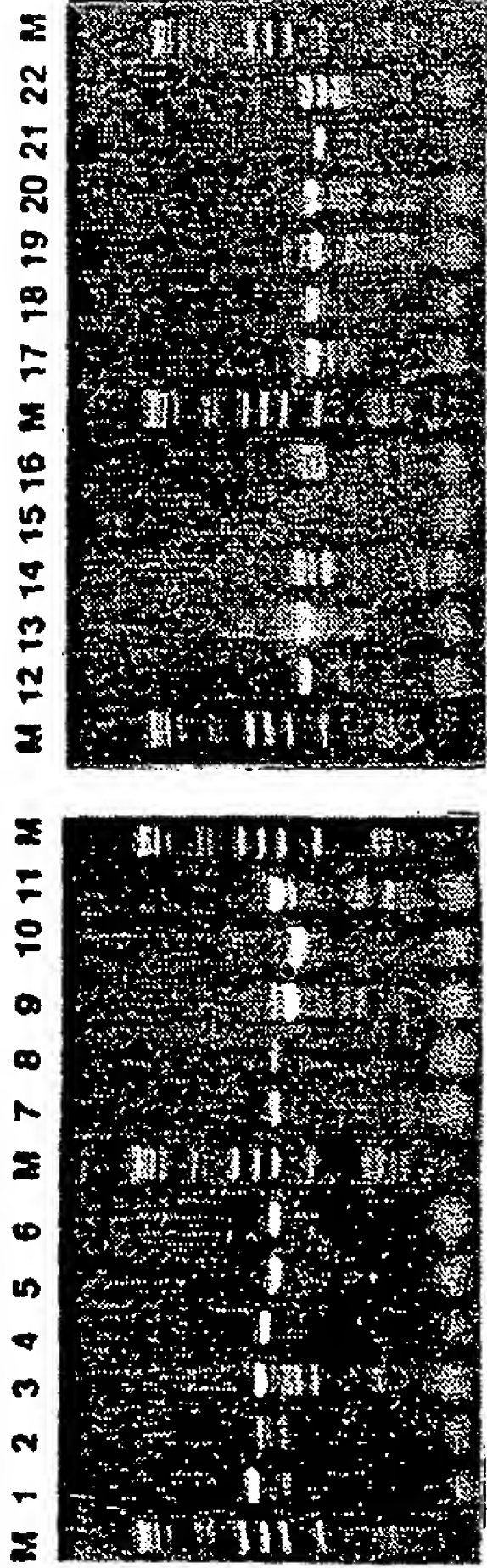
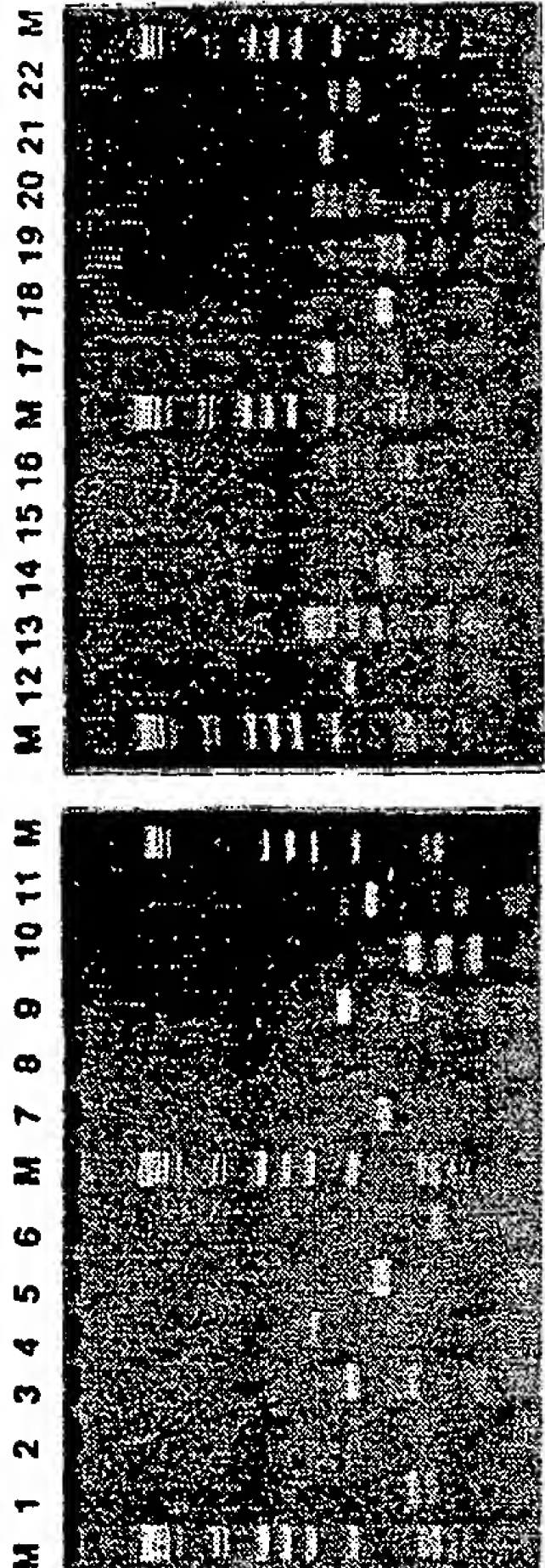
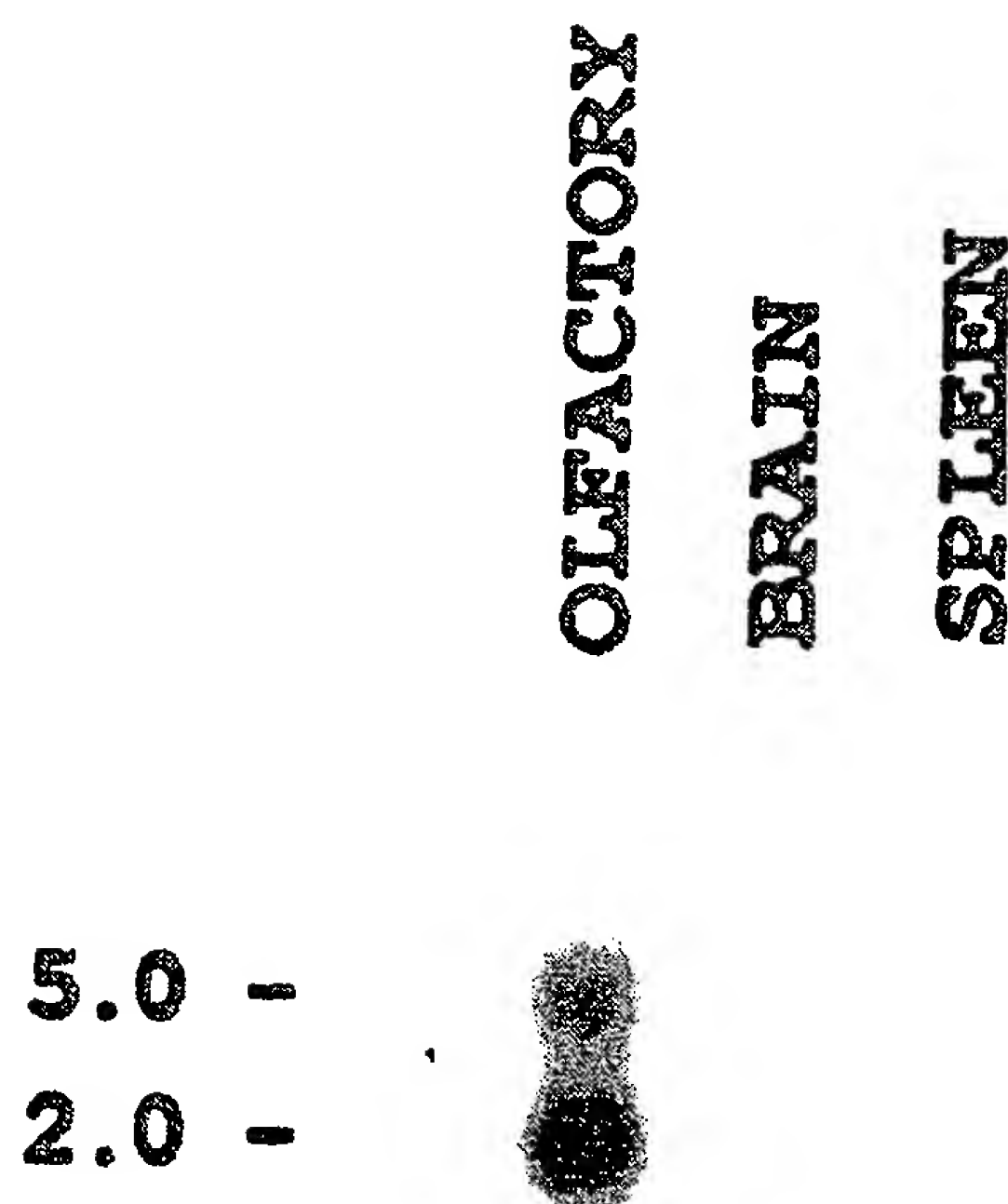


Figure 2B



4/99

Figure 3



5/99

Figur 4A

F3		N	D	S	S	N	R	T	R	V	S	E	11		
F5		N	S	S	T	N	Q	S	S	V	T	E	11		
F6	N	A	W	S	T	G	Q	N	L	S	T	P	G	P	14
F12		N	E	S	G	N	S	T	R	R	F	S	S	12	
I3		N	N	-	-	N	Q	T	F	I	T	Q	9		
I7		N	E	R	R	N	H	S	G	R	V	S	E	12	
I8		N	N	-	-	N	K	T	V	I	T	H	9		
I8		N	T	R	R	N	Q	T	A	I	S	Q	11		
I14		N	T	G	N	N	Q	T	L	I	L	E	11		
I15		N	T	E	E	N	Q	T	V	I	S	Q	11		

F3	F	L	L	L	G	F	V	E	N	K	D	L	Q	P	25
F5	F	L	L	L	G	L	S	R	Q	P	Q	Q	Q	25	
F6	F	I	L	L	G	F	P	G	P	R	S	M	R	I	28
F12	F	F	L	L	G	F	T	E	N	P	Q	L	H	F	26
I3	F	L	L	L	G	L	P	I	P	E	E	H	Q	H	23
I7	F	V	L	L	G	F	P	A	P	A	P	L	R	V	26
I8	F	L	L	L	G	L	P	I	P	P	E	H	Q	Q	23
I9	F	F	L	L	G	L	P	F	P	P	E	Y	Q	H	25
I14	F	L	L	L	G	L	P	I	P	S	E	Y	H	L	25
I15	F	L	L	L	F	L	P	I	P	S	E	H	Q	H	25

6/99

Figur 4B

	<u>I</u>	
F3	L I Y G L F L S N Y L V T V	39
F5	L L F L L F L I N Y L A T V	39
F6	G L F L L F L V N Y L L T V	42
F12	L I F A L F L S N Y L V T V	40
I3	L F Y A L F L V N Y L T T I	37
I7	L L F F L S L L X Y V L V L	40
I8	L F F A L F L I N Y L T T F	37
I9	L F Y A L F L A N Y L T T L	39
I14	L F Y A L F L A N Y L T I I	29
I15	V F Y A L F L S N Y L T T V	39

	<u>I</u>	
F3	I G N I S I I V A I I S D P	53
F5	L G N L L I I L A I G T D S	53
F6	V G N L A I I S L V G A H R	56
F12	L G N L L I I M A I I T Q S	54
I3	L G N L L I I V L V Q L D S	51
I7	T E N M L I I I A I R N H P	54
I8	L G N L L I V V L V Q L D S	51
I9	L G N L I I I I L I L L D S	53
I14	L G N L L I I V L V R L D S	53
I15	L G N L I I I I L I H L D S	53

7/99

Figure 4C

	<u>II</u>														
F3	C	L	H	T	P	M	Y	F	F	L	S	N	L	S	67
F5	R	L	H	T	P	N	Y	F	F	L	S	N	L	S	67
F6	C	L	Q	T	P	M	Y	F	F	L	C	N	L	S	70
F12	H	L	H	T	P	N	Y	F	F	L	A	N	L	S	68
I3	Q	L	H	T	P	N	Y	L	F	L	S	N	L	S	65
I7	T	L	H	K	P	M	Y	F	F	L	A	N	M	S	68
I8	H	L	H	T	P	M	Y	L	F	L	S	N	L	S	65
I9	H	L	H	T	P	N	Y	L	F	L	S	N	L	S	67
I14	H	L	H	M	P	N	Y	L	F	L	S	N	L	S	67
I15	H	L	H	T	P	M	Y	L	F	L	S	N	L	S	67

	<u>II</u>														
F3	F	V	D	I	C	F	I	S	T	T	V	P	K	M	81
F5	F	V	D	V	C	F	S	S	T	T	V	P	K	V	81
F6	F	L	E	I	W	F	T	T	A	C	V	P	K	T	84
F12	F	V	D	I	C	F	T	S	T	T	I	P	K	M	82
I3	F	S	D	L	C	F	S	S	V	T	M	P	K	L	79
I7	F	L	E	I	W	Y	V	T	V	T	I	P	K	M	82
I8	F	S	D	L	C	F	S	S	V	T	M	L	K	L	79
I9	F	A	D	L	C	F	S	S	V	T	M	P	K	L	67
I14	F	S	D	L	C	F	S	S	V	T	M	P	K	L	67
I15	F	S	D	L	C	F	S	S	V	T	M	P	K	L	67

9/99

Figure 4E

	<u>III</u>	
F3	L F V E L D N F L L T I N A	119
F5	V F G N M D N F L L A V N S	119
F6	S L G C T E Y F L L A V N A	122
F12	V F A I L G N F L L A V N A	120
I3	V F G D M E S F L L V A N A	117
I7	G L G C T E C V L L A V N A	124
I8	L F G Y L G N F L L V A N A	117
I9	F F G D L G N F L L V A N A	119
I14	V F G D M E S F L L V V N A	119
I15	Y F A D L E S F L L V A N A	119

	<u>III</u>	
F3	Y D R Y V A I C H P M H Y T	133
F5	Y D R F V A I C H P L H Y T	133
F6	Y D R Y L A I C L P L R Y G	136
F12	Y D R Y V A X C H P L C Y T	134
I3	Y D R Y V A I C F P L H Y T	131
I7	Y D R Y V A I C H P L H Y P	138
I8	Y D R Y V A I C F P L H Y T	131
I9	Y D R Y V A I C F P L H Y M	133
I14	Y D R Y V A I C F P L R Y T	133
I15	Y D R Y V A I C F P L H Y M	133

10/99

Figure 4F

	<u>IV</u>														
F3	V	I	N	N	Y	K	L	C	G	F	L	V	L	V	147
F5	T	K	N	T	R	Q	L	C	V	L	L	V	V	G	147
F6	G	I	N	T	P	G	L	A	M	R	L	A	L	G	150
F12	V	I	V	N	H	R	L	C	I	L	L	L	L	L	148
I3	S	I	N	S	P	K	L	C	T	C	L	V	L	L	145
I7	V	I	V	S	S	R	L	C	V	Q	M	A	A	G	152
I8	N	I	N	S	H	K	L	C	T	C	L	L	L	V	145
I9	S	I	N	S	P	K	L	C	V	S	L	V	V	L	147
I14	T	I	N	S	T	K	F	C	A	S	L	V	L	L	147
I15	S	I	N	S	P	K	L	C	V	S	L	V	V	L	147

	<u>IV</u>														
F3	S	W	I	V	S	V	L	H	A	L	F	Q	S	L	161
F5	S	W	V	V	A	N	M	N	C	L	L	H	I	L	161
F6	S	W	L	C	G	F	S	A	I	T	V	P	A	T	164
F12	S	W	V	I	S	I	F	H	A	F	I	Q	S	L	162
I3	L	W	M	L	T	T	S	H	A	M	M	H	T	L	159
I7	S	W	A	G	G	F	G	I	S	M	V	K	V	F	166
I8	F	W	I	M	T	S	S	H	A	M	M	H	T	L	159
I9	S	W	V	L	T	T	F	H	A	M	L	H	T	L	161
I14	L	W	M	L	T	M	T	H	A	L	L	H	T	L	161
I15	S	W	V	L	T	T	F	H	A	M	L	H	T	L	161

11/99

Figur 4G

F3	M	M	L	A	L	P	F	C	T	H	L	E	I	P	175
F5	L	M	A	R	K	S	F	C	A	D	N	M	I	P	175
F6	L	I	A	R	L	S	F	C	G	S	R	V	I	N	178
F12	I	V	L	Q	L	T	F	C	G	D	V	K	I	P	176
I3	L	A	A	R	L	S	F	C	E	N	N	V	V	L	173
I7	L	I	S	R	L	S	Y	C	G	P	N	T	I	N	180
I8	L	A	A	R	L	S	F	C	E	N	N	V	L	L	173
I9	L	M	A	R	L	S	F	C	E	D	S	V	I	P	175
I14	L	I	A	R	L	S	F	C	E	K	N	V	I	L	175
I15	L	M	A	R	L	S	F	C	A	D	N	M	I	P	175

F3	H	Y	F	C	E	P	N	Q	V	I	Q	L	T	C	189
F5	H	F	F	C	D	G	T	P	L	L	K	L	S	C	189
F6	H	F	F	C	D	I	S	P	W	I	V	L	S	C	192
F12	H	F	F	C	E	L	N	Q	L	S	Q	L	T	C	190
I3	N	F	F	C	D	L	F	V	L	L	K	L	A	C	187
I7	H	F	F	C	D	V	S	P	L	L	N	L	S	C	194
I8	N	F	F	C	D	L	F	V	L	L	K	L	A	C	187
I9	H	Y	F	C	D	M	S	T	L	L	K	V	A	C	189
I14	H	F	F	C	D	I	S	A	L	L	K	L	S	C	189
I15	H	F	F	C	D	I	S	P	L	L	K	L	S	C	189

12/99

Figur 4H

	<u>V</u>														
F3	S	D	A	F	L	N	D	L	V	I	Y	F	T	L	203
F5	S	D	T	H	L	N	E	L	M	I	L	T	E	G	203
F6	T	D	T	Q	V	V	E	L	V	S	F	G	I	A	206
F12	S	D	N	F	P	S	H	L	I	M	N	L	V	P	204
I3	S	D	T	Y	I	N	E	L	M	I	F	I	M	S	201
I7	T	D	M	S	T	A	E	L	T	D	F	V	L	A	208
I8	S	D	T	Y	V	N	E	L	M	I	H	I	M	G	201
I9	S	D	T	H	D	N	E	L	A	I	F	I	L	G	203
I14	S	D	I	Y	V	N	E	L	M	I	Y	I	L	G	203
I15	S	D	T	H	V	N	E	L	V	I	F	V	M	G	203

	<u>V</u>														
F3	V	L	L	A	T	V	P	L	A	G	I	F	Y	S	217
F5	A	V	V	M	V	T	P	F	V	C	I	L	I	S	217
F6	F	C	V	I	L	G	S	C	G	I	T	L	V	S	220
F12	V	M	L	A	A	I	S	F	S	G	I	L	Y	S	218
I3	T	L	L	I	I	I	P	F	F	L	I	V	M	S	215
I7	I	F	I	L	L	G	P	L	S	V	T	G	A	S	222
I8	V	I	I	I	V	I	P	F	V	L	I	V	I	S	215
I9	G	P	I	V	V	L	P	F	L	L	I	I	V	S	203
I14	G	L	I	I	I	I	P	F	L	L	I	V	M	S	203
I15	G	L	V	I	V	I	P	F	V	L	I	I	V	S	203

13/99

Figure 4I

	<u>V</u>	
F3	Y F K I V S S I C A I S S V	231
F5	Y I H I T C A V L R V S S P	231
F6	Y A Y I I T T I I K I P S A	234
F12	Y F K I V S S I H S I S T V	232
I3	Y A R I I S S I L K V P S T	229
I7	Y M A I T G A V M R I P S A	236
I8	Y A K I I S S I L K V P S T	229
I9	Y A R I V S S I F K V P S S	231
I14	Y V R I F F S I L K F P S I	231
I15	Y A R V V A S I L K V P S V	231

	<u>VI</u>	
F3	H G K Y K A F S T C A S H L	245
F5	R G G W K S F S T C G S H L	245
F6	R G R H R A F S T C S S H L	248
F12	Q G K Y K A F S T C A S H L	246
I3	Q G I C K V F S T C G S H L	243
I7	A G R H K A F S T C A S H L	250
I8	Q S I H K V F S T C G S H L	243
I9	Q S I H K A F S T C G S H L	245
I14	Q D I Y K V F S T C G S H L	245
I15	R G I H K I F S T C G S H L	245

14/99
Figure 4J

	<u>VI</u>	
F3	S V V S L F Y C T G L G V Y	259
F5	A V V C L F Y G T V I A V Y	259
F6	T V V L I W Y G S T I F L H	262
F12	S I V S L F Y S T G L G V Y	260
I3	S V V S L F Y G T I I G L Y	257
I7	T V V I I F Y A A S I F I Y	264
I8	S V V S L F Y G T I I G L Y	257
I9	S V V S L F Y G T V I G L Y	259
I14	S V V T L F Y G T I F G I Y	259
I15	S V V S L F Y G T I I G L Y	259

	<u>VI</u>	<u>VII</u>	
F3	L S S A A N N S S Q A S A T		273
F5	F N P S S S H L A G R D M A		273
F6	V R T S V E S S L D L T K A		276
F12	V S S A V V Q S S H S A A S		274
I3	L C P A G N N S T V K E M V		271
I7	A R P K A L S A F D T N K L		278
I8	L C P S G D N F S L K G S A		271
I9	L C P S A N N S T V K E T V		273
I14	L C P S G N N S T V K E I A		273
I15	L C P S A N N S T V K E T V		273

15/99

Figur 4K

	<u>VII</u>														
F3	A	S	V	M	Y	T	V	V	T	P	M	V	N	P	287
F5	A	A	V	M	Y	A	V	V	T	P	M	L	N	P	287
F6	I	T	V	L	N	T	I	V	T	P	V	L	N	P	290
F12	A	S	V	M	Y	T	V	V	T	P	M	L	N	P	288
I3	M	A	M	M	Y	T	V	V	T	P	M	L	N	P	285
I7	V	S	V	L	Y	A	V	I	V	P	L	F	N	P	292
I8	M	A	M	M	Y	T	V	V	T	P	M	L	N	P	285
I9	M	S	L	M	Y	T	M	V	T	P	M	L	N	P	287
I14	M	A	M	M	Y	T	V	V	T	P	M	L	N	P	287
I15	M	A	M	M	Y	T	V	V	T	P	M	L	N	P	287

	<u>VII</u>														
F3	F	I	Y	S	L	R	N	K	D	V	K	S	V	L	301
F5	F	I	Y	S	L	R	N	S	D	M	K	A	A	L	301
F6	F	I	Y	T	L	R	N	K	D	V	K	E	A	L	304
F12	F	I	Y	S	L	R	N	K	D	V	K	R	A	L	302
I3	F	I	Y	S	L	R	N	R	D	M	K	R	A	L	299
I7	I	I	Y	C	L	R	N	Q	D	V	K	R	A	L	306
I8	F	I	Y	S	L	R	N	R	D	M	K	Q	A	L	299
I9	F	I	Y	S	L	R	N	R	D	I	K	D	A	L	301
I14	F	I	Y	S	L	R	N	R	D	M	K	R	A	L	301
I15	F	I	Y	S	L	R	N	R	D	M	K	E	A	L	301

16/99

Figure 4L

F3	K K T L C E E V I R S P P S	315
F5	R K V L A M R F P S K Q -	313
F6	R R T V K G K -	311
F12	E R L L E G N C K V H H W T	316
I3	I R V I C S M K I T L -	310
I7	R R T L H L A Q D Q E A N T	320
I8	I R V T C S K K I S L P W -	312
I9	E K I M C K K Q I P S F L -	314
I14	I R V I C T K K I S L -	312
I15	I R V L C K K K I T F C L -	314

F3	L L H F F L V L C H L P C F	329
F5		
F6		
F12	G -	317
I3		
I7	N K G S K I G -	327
I8		
I9		
I14		
I15		

17/99

Figur 4M

F3
F5
F6
F12
I3
I7
I8
I9
I14
I15

I F C Y -

333

18/99

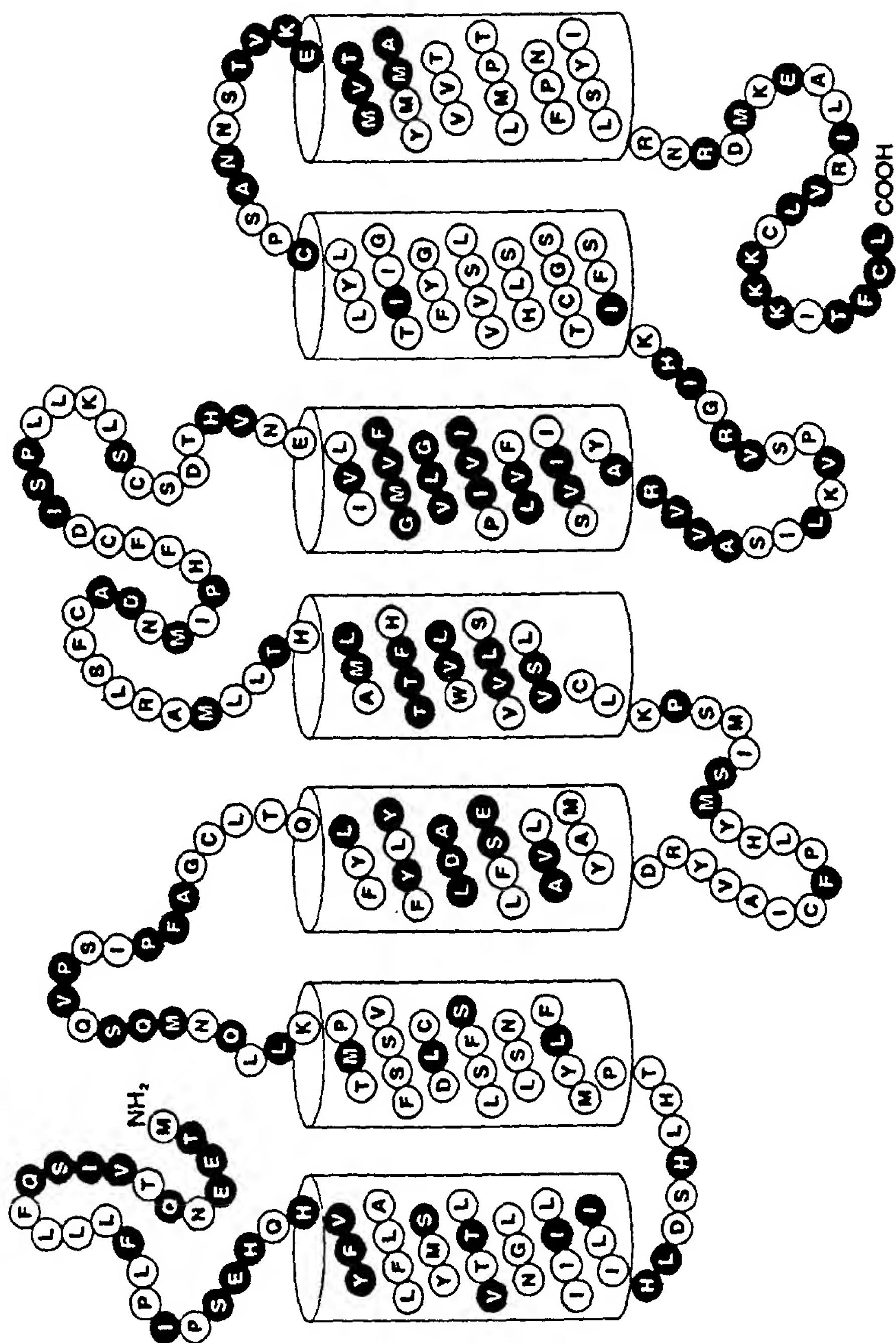


Figure 5

19/99

Figure 6A(1)

					V									
F2	R	V	N	E	V	V	I	F	I	V	V	S	L	F
F3	F	L	N	D	L	V	I	Y	F	T	L	V	L	L
F5	H	L	N	E	L	M	I	L	T	E	G	A	V	V
F6	Q	V	V	E	L	V	S	F	G	I	A	F	C	V
F7	H	V	N	E	L	V	I	F	V	M	G	G	I	I
F8	F	P	S	H	L	T	M	H	L	V	P	V	I	L
F12	F	P	S	H	L	I	M	N	L	V	P	V	M	L
F13	F	P	S	H	L	I	M	N	L	V	P	V	M	L
F23	F	L	N	D	V	I	M	Y	F	A	L	V	L	L
F24	H	E	I	E	M	I	I	L	V	L	A	A	F	N
I3	Y	I	N	E	L	M	I	F	I	M	S	T	L	L
I7	S	T	A	E	L	T	D	F	V	L	A	I	F	I
I8	Y	V	N	E	L	M	I	H	I	M	G	V	I	I
I9	H	D	N	E	L	A	I	F	I	L	G	G	P	I
I11	H	L	N	E	L	M	I	L	T	E	G	A	V	V
I12	F	P	S	H	L	I	M	N	L	V	P	V	M	L
I14	Y	V	N	E	L	M	I	Y	I	L	G	G	L	I
I15	H	V	N	E	L	V	I	F	V	M	G	G	L	V

20/99

Figure 6A(2)

	<u>V</u>												
F2	L	V	L	P	F	A	L	I	I	M	S	Y	V R
F3	A	T	V	P	L	A	G	I	F	Y	S	Y	F K
F5	M	V	T	P	F	V	C	I	L	I	S	Y	I H
F6	I	H	G	S	C	G	I	T	L	V	S	Y	A Y
F7	L	V	I	P	F	V	L	I	I	V	S	Y	V R
F8	A	A	I	S	L	S	G	I	L	Y	S	Y	F K
F12	A	A	I	S	F	S	G	I	L	Y	S	Y	F K
F13	A	A	I	S	F	S	G	I	L	Y	S	Y	F K
F23	A	V	V	P	L	L	G	I	L	Y	S	Y	S K
F24	L	I	S	S	L	L	V	V	L	V	S	Y	L F
I3	I	I	I	P	F	F	L	I	V	M	S	Y	A R
I7	L	L	G	P	L	S	V	T	G	A	S	Y	M A
I8	I	V	I	P	F	V	L	I	V	I	S	Y	A K
I9	V	V	L	P	F	L	L	I	I	V	S	Y	A R
I11	M	V	T	P	F	V	C	I	L	I	S	Y	I H
I12	G	A	I	S	L	S	G	I	L	Y	S	Y	F K
I14	I	I	I	P	F	L	L	I	V	M	S	Y	V R
I15	I	V	I	P	F	V	L	I	I	V	S	Y	A R

21/99

Figure 6A(3)

F2	I	V	S	S	I	L	K	V	P	S	S	Q	G	I
F3	I	V	S	S	I	C	A	I	S	S	V	H	G	K
F5	I	T	C	A	V	L	R	V	S	S	P	R	G	G
F6	I	I	T	T	I	I	K	I	P	S	A	R	G	R
F7	I	V	S	S	I	L	K	V	P	S	A	R	G	I
F8	I	V	S	S	I	R	S	M	S	S	V	Q	G	K
F12	I	V	S	S	I	H	S	I	S	T	V	Q	G	K
F13	I	V	S	S	I	R	S	V	S	S	V	K	G	K
F23	I	V	S	S	I	R	A	I	S	T	V	Q	G	K
F24	I	L	I	A	I	L	R	M	N	S	A	E	G	R
I3	I	I	S	S	I	L	K	V	P	S	T	Q	G	I
I7	I	T	G	A	V	M	R	I	P	S	A	A	G	R
I8	I	I	S	S	I	L	K	V	P	S	T	Q	S	I
I9	I	V	S	S	I	F	K	V	P	S	S	Q	S	I
I11	I	T	W	A	V	L	R	V	S	S	P	R	G	G
I12	I	V	S	S	V	R	S	I	S	S	V	Q	G	K
I14	I	F	F	S	I	L	K	F	P	S	I	Z	D	I
I15	V	V	A	S	I	L	K	V	P	S	V	R	G	I

22/99

Figure 6A(4)

F2	Y	K
F3	Y	K
F5	W	K
F6	H	R
F7	R	K
F8	Y	K
F12	Y	K
F13	Y	K
F23	Y	K
F24	R	K
I3	C	K
I7	H	K
I8	H	K
I9	H	K
I11	W	K
I12	H	K
I14	Y	K
I15	H	K

23/99

Figure 6B

					<u>V</u>								
F12	F	P	S	H	L	I	M	N	L	V	P	V	M
F13	F	P	S	H	L	I	M	N	L	V	P	V	M
F8	F	P	S	H	L	T	M	H	L	V	P	V	I
I12	F	P	S	H	L	I	M	N	L	V	P	V	M
F23	F	L	N	D	V	I	M	Y	F	A	L	V	L
F3	F	L	N	D	L	V	I	Y	F	T	L	V	L

					<u>V</u>								
F12	A	A	I	S	F	S	G	I	L	Y	S	Y	F
F13	A	A	I	S	F	S	G	I	L	Y	S	Y	F
F8	A	A	I	S	L	S	G	I	L	Y	S	Y	F
I12	G	A	I	S	L	S	G	I	L	Y	S	Y	F
F23	A	V	V	P	L	L	G	I	L	Y	S	Y	S
F3	A	T	V	P	L	A	G	I	F	Y	S	Y	F

24/99

Figure 6B (Continued)

F12	I	V	S	S	I	H	S	I	S	T	V	Q	G	K
F13	I	V	S	S	I	R	S	V	S	S	V	K	G	K
F8	I	V	S	S	I	R	S	M	S	S	V	Q	G	K
I12	I	V	S	S	V	R	S	I	S	S	V	Q	G	K
F23	I	V	S	S	I	R	A	I	S	T	V	Q	G	K
F3	I	V	S	S	I	C	A	I	S	S	S	H	G	K

F12	Y	K
F13	Y	K
F8	Y	K
I12	H	K
F23	Y	K
F3	Y	K

25/99

Figure 6C

					V								
F7	H	V	N	E	L	V	I	F	V	M	G	G	I
I15	H	V	N	E	L	V	I	F	V	M	G	G	L
I3	Y	I	N	E	L	M	I	F	I	M	S	T	L
I8	Y	V	N	E	L	M	I	H	I	M	G	V	I
I9	H	D	N	E	L	A	I	F	I	L	G	G	P
I14	Y	V	N	E	L	M	I	Y	I	L	G	G	L

					V								
F7	L	V	I	P	F	V	L	I	I	V	S	Y	V
I15	I	V	I	P	F	V	L	I	I	V	S	Y	A
I3	I	I	I	P	F	F	L	I	V	M	S	Y	A
I8	I	V	I	P	F	V	L	I	V	I	S	Y	A
I9	V	V	L	P	F	L	L	I	I	V	S	Y	A
I14	I	I	I	P	F	L	L	I	V	M	S	Y	V

Figure 6C (Continued)

F7	I	V	S	S	I	L	K	V	P	S	A	R	G	I
I15	V	V	A	S	I	L	K	V	P	S	V	R	G	I
I3	I	I	S	S	I	L	K	V	P	S	T	Q	G	I
I8	I	I	S	S	I	L	K	V	P	S	T	Q	S	I
I9	I	V	S	S	I	F	K	V	P	S	S	Q	S	I
I14	I	F	F	S	I	L	K	F	P	S	I	Q	D	I

F7	R	K
I15	H	K
I3	C	K
I8	H	K
I9	H	K
I14	Y	K

27/99

Figure 6D

					V											
F5	H	L	N	E	L	M	I	L	T	E	G	A	V	V		
I11	H	L	N	E	L	M	I	L	T	E	G	A	V	V		

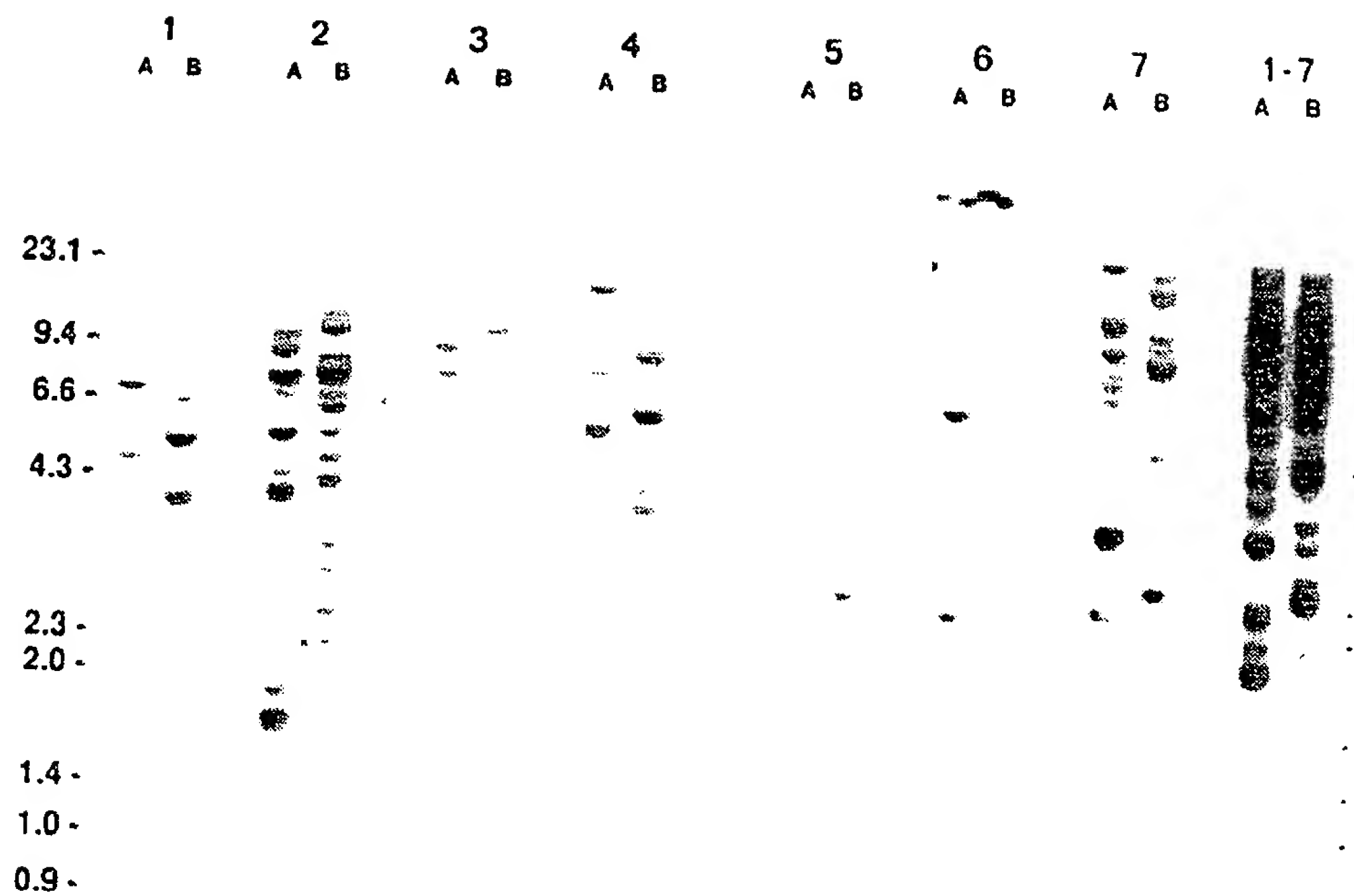
	V																		
F5	N	V	T	P	F	V	C	I	L	I	S	Y	I	H					
I11	N	V	T	P	F	V	C	I	L	I	S	Y	I	H					

F5 I T C A V L R V S S P R G G
I11 I T W A V L R V S S P R G G

F5	W	K
I11	W	K

28/99

Figure 7



29/99

Figure 8

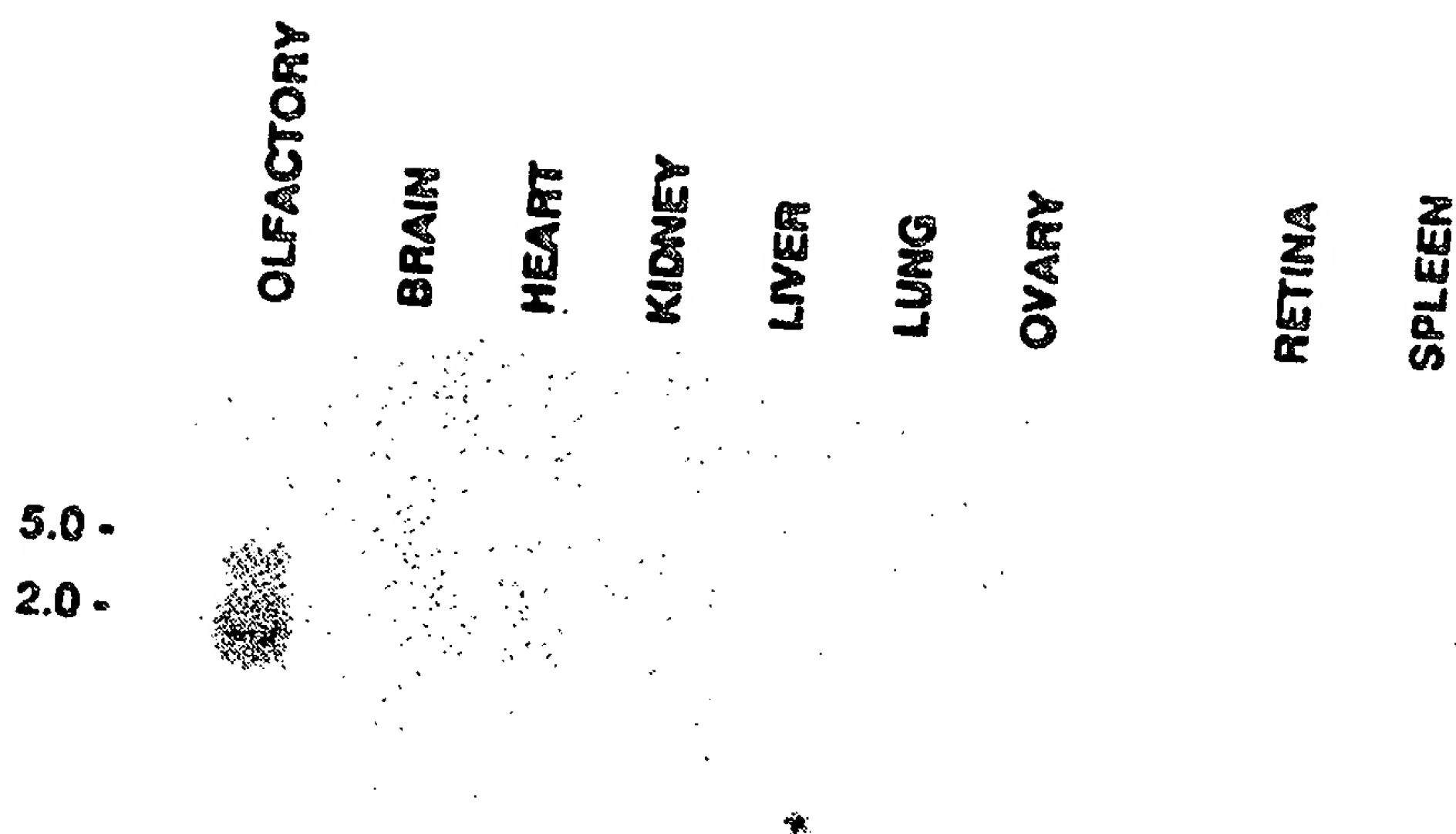


Figure 9A Translated sequence of F3T.D1S

10	20	30	40	50	60
* ATC GAC TCA AGC AAC AGG ACA AGA GTT TCA GAA TTT CTT CGA TTT GTA GAA AAC M D S S N R T R V S E F L L C F V E N	* 20	* 30	* 40	* 50	* 60
70	80	90	100	110	120
* AAA GAC CTA CAA CCC CTT ATT TAT GGT CTT TTT CTC TCT ATG TAC CTG GTT ACT GTC ATT K D L Q P L I Y G L F L S M Y L V T V I	* 80	* 90	* 100	* 110	* 120
130	140	150	160	170	180
* GGA AAC ATA TCC ATT ATT GTG GCT ATC ATT TCA GAT CCC TGT CTG CAC ACC CCC ATG TAT G N I S I I V A I I S D P C L H T P M Y	* 140	* 150	* 160	* 170	* 180
190	200	210	220	230	240
* TTC TTC CTC TCT AAC CTG TCC TTT GTG GAC ATC TGT TTC ATT TCA ACC ACT GTT CCA AAC F L S N L S F V D I C F I S T T V P K	* 200	* 210	* 220	* 230	* 240
250	260	270	280	290	300
* ATC TTA GTG AAC ATC CAG ACC CAA AAC AAT GTC ATC ACC TAT GCA GGA TGC ATT ACC CAG M L V N I Q T Q N V I T Y A C I T Q	* 260	* 270	* 280	* 290	* 300

30/99

Figure 9B

310 * 320 * 330 * 340 * 350 * 360 *
ATA TAC TTT TTC TTG CTC TTT GTA GAA TTG GAC AAC TTC TTG CTG ACT ATC ATG GCC TAT
I Y F F L L L F V E L D N F L L T I M A Y

370 * 380 * 390 * 400 * 410 * 420 *
GAC CGT TAC GTA GCC ATC TGT CAC CCC ATG CAC TAC ACA GTT ATC ATG AAC TAC AAC CTC
D R Y V A I C H P M H Y T V I M N Y K L

430 * 440 * 450 * 460 * 470 * 480 *
TGT GCA TTT CTG GTT CTG GTA TCT TGG ATT GTA AGT GTT CTG CAT GCC TTG TTT CAA ACC
C G F L L V L V S S W I V S V L L H A L F Q S

490 * 500 * 510 * 520 * 530 * 540 *
TTG ATG ATG TTG GCG CTG CCC TTC TGC AGA CAT CTG GAA ATC CCA CAC TAC TTC TGT GAA
L M M L L A L P F C T H L E I P H Y F C E

550 * 560 * 570 * 580 * 590 * 600 *
CCT AAT CAG GTG ATT CAA CTC ACC TGT TCT GAT GCA TTT CTT AAT GAT CTT GTG ATA TAT
P N Q V I Q L T C S D A F L L N D L V I Y

610 620 630 640 650 660

31/99

[illegible]

33/99

Figure 9D

CTG AAA AAA ACT CTT TGT GAG GAA GTT ATA AGG AGT CCA CCT TCC CTA CTT CAT TTC TTC	910	920	930	940	950	960
L K T L C E E V I R S P P P L L H F F	*	*	*	*	*	*
CTA GTG TTA TGT CAT CTC CCT TGT TTT ATT TTT TGT TAT TAA	970	980	990	1000		
L V L C H L P C F I F C Y -	*	*	*	*		

Translation begun with base no. 57

Translated to base no.1058

Sequence printed from base no. 57 to base no.1058

Sequence numbered beginning with base no. 57

ה'תשנ"ח

[illegible]

35/99

Figure 10B

CTG TAT TTT CTC GCT GTG TTT GGT AAC ATG GAC AAT TTC CTG GCT CTG ATG TCC TAT	310	320	330	340	350	360
L Y F L A V F G N M D N F L L A V M S Y	*	*	*	*	*	*
GAC CGA TTT GTG GCC ATA TGC CAC CCT TTA CAC TAC ACA AAG ATG ACC CGT CAG CTC	370	380	390	400	410	420
D R F V A I C H P L H Y T T K M T R Q L	*	*	*	*	*	*
IGT CTC CTG CTT GTT GTG GGG TCA TGG GTT GTA GCC AAC ATG AAT TGT CTG TTG CAC ATA	430	440	450	460	470	480
C V L L V V G S W V V A N M N C L L H I	*	*	*	*	*	*
CTG CTC ATG GCT CGA CTC TCC TCC TTC TGT GCA GAC AAC ATG ATG CCC CAC TTC TGT GAT	490	500	510	520	530	540
L L M A R L L S F C A D N M I P H F C D	*	*	*	*	*	*
GGA ACT CCC CTC CTG AAA CTC TCC TGC TCA GAC ACA CAT CTC AAT GAG CTG ATG ATT CTT	550	560	570	580	590	600
G T P L L K L S C S D T H L N E L M I L	*	*	*	*	*	*
	610	620	630	640	650	660

Figure 10C

ACA	GAG	GGA	GCT	GTC	GTC	ATG	GTC	ACC	CCA	TTT	GTC	TGC	ATC	CTC	ATC	TCC	TAC	ATC	CAC	*
T	E	G	A	V	V	M	V	T	P	F	V	C	I	L	I	S	Y	I	H	*
670						680			690			700				710			720	
*						*			*			*				*			*	
ATC	ACC	TGT	GCT	GTC	CTC	AGA	GTC	TCA	TCC	CCC	AGG	GGA	GGA	TGG	AAA	TCC	TTC	TCC	ACC	
I	T	C	A	V	L	R	V	S	S	P	R	G	G	W	K	S	F	S	T	
730						740			750			760				770			780	
*						*			*			*				*			*	
TGT	GGC	TCC	CAC	CTG	GCT	GTC	GTC	TGC	CTC	TTC	TAT	GGC	ACC	GTC	ATC	GCT	GTG	TAT	TTC	
C	G	S	H	L	A	V	V	C	L	F	Y	G	T	V	I	A	V	Y	F	
790						800			810			820				830			840	
*						*			*			*				*			*	
AAC	CCA	TCA	TCC	TCT	CAC	TTA	GCT	GGG	AGG	GAC	ATG	GCA	GCT	GCA	GTG	ATG	TAT	GCA	GTG	
PRONUC/TRA OPTION																				
N	P	S	S	S	H	L	A	C	R	D	M	A	A	A	V	M	Y	A	V	

37/99

Figure 10D

GTG ACC CCA ATG CTG AAC CCT TTC ATC TAT AGC CTG AGG AAC AGC GAC ATG AAA GCA GCT	850	860	870	880	890	900
V T P M L N P F I Y S L R N M K A A	*	*	*	*	*	*
TTA AGG AAA GTG CTC GCC ATG AGA TTT CCA TCT AAG CAG TAA	910	920	930	940		
L R K V L A M R F P S K Q -	*	*	*	*		

Translation begun with base no. 62
 Translated to base no.1003
 Sequence printed from base no. 62 to base no.1003
 Sequence numbered beginning with base no. 62

Figure 11A Translated sequence of F6T.D1S

10 * 20 * 30 * 40 * 50 * 60 *
 ATG GCT TCG AGT ACT GGC CAG AAC CTG TCC ACA CCA GGA CCA TTC ATC TTC CTG GGC TTC
 M A W S T G C Q N L S T P G P F I L L C F

 70 * 80 * 90 * 100 * 110 * 120 *
 CCA GGG CCA AGG AGC ATG CGC ATT GGG CTC TTC CTG CTT TTC CTG GTC ATG TAT CTG CTT
 P G P R S M R I G L F L L F L V M Y L L

 130 * 140 * 150 * 160 * 170 * 180 *
 ACG GTA GTT GGA AAC CTA GCC ATC ATC TCC CTG GTA GGT GCC CAC AGA TGC CTA CAG ACA
 T V V G C N L A I I S L V G A H R C L Q T

 190 * 200 * 210 * 220 * 230 * 240 *
 CCC ATG TAC TTC CTC TGC AAC CTC TCC TTC CTG GAG ATC TGG TTC ACC ACA GCC TGC
 P M Y F F L C N L S F L E I W F T A C

 250 * 260 * 270 * 280 * 290 * 300 *
 GTA CCC AAG ACC CTG GCC ACA TTT GCG CCT CGG GGT GGA CTC ATT TCC TTG CCT GCC TGT
 V P K T L A T F A P R G G V I S L A G C

38/99

Figure 11B

310 * 320 * 330 * 340 * 350 * 360 *
 GCC ACA CAG ATG TAC TTT CTC TTT TCT TTG GGC TGT ACC GAG TAC TTC CTG GCT CTG
 A T Q M Y F V F S L G C T E Y F L A V

 370 * 380 * 390 * 400 * 410 * 420 *
 ATG GCT TAT GAC CGC TAC CTG GGC ATC TGC CTG CCA CTG CGC TAT GGT GGC ATC ATG ACT
 M A Y D R Y L A I C L P L R Y G G I M T

 430 * 440 * 450 * 460 * 470 * 480 *
 CCT GGG CTG GCG ATG CGG TTG GGC GGA TCC TGG CTG TGT GGT TTT TCT GCA ATC ACA
 P G L A M R L L A L G S W L C G F S A I T
 39/99

 490 * 500 * 510 * 520 * 530 * 540 *
 GTT CCT GCT ACC CTC ATT CCC GGC CTC TCT TTC TGT GGC TCA CGT GTC ATC AAC CAC TTC
 V P A T L I A R L S F C G S R V I N H F

 550 * 560 * 570 * 580 * 590 * 600 *
 TTC TGT GAC ATT TCG CCC TGG ATA GTG CTT TCC TGC ACC GAC ACC CAG CTG GAA CTG
 F C D I S P W I V L S C T D T Q V E L

 610 620 630 640 650 660

Figure 11c

GTG	TCC	TTT	GGC	ATT	GGC	TTC	TGT	GTT	ATT	CTG	GGC	TCG	TGT	GGT	ATC	ACA	CTA	GTC	TCC	*
V	S	F	G	I	A	F	C	V	I	L	G	S	C	G	I	T	L	V	S	*
670																				720
TAT	GCT	TAC	ATC	ATC	ACT	ACC	ATC	ATC	AAG	ATT	CCC	TCT	GCC	CGG	GGC	CGG	CAC	CGC	GCC	*
Y	A	Y	I	I	T	T	I	I	K	I	P	S	A	R	G	R	H	R	A	*
730																				780
TTC	TCA	ACC	TGC	TCA	TCC	CAT	CTC	ACT	GTG	GTG	CTG	ATT	TGG	TAT	GGC	TCC	ACC	ATC	TTC	40/99
F	S	T	C	S	S	H	L	T	V	V	L	I	W	Y	G	S	T	I	F	*
790																				840
TTG	CAT	GTG	AGG	ACC	TCG	GTA	GAG	AGC	TCC	TTG	GAC	CTC	ACC	AAA	GCT	ATC	ACA	GTG	CTG	*
PRONUC/TRA OPTION																				
L	H	V	R	T	S	V	E	S	S	L	D	L	T	K	A	I	T	V	L	

Figure 11D

	850		860		870		880		890		900
	*		*		*		*		*		*
AAC ACC ATT GTC ACA CCT GTG CTG AAC CCT TTC ATA TAT ACT CTG AGG AAC AAG GAT CTC											
N T I V T P V L N P F I Y T L R N K D V											

	910		920		930
*		*		*	
AAG GAA GCT CTG CGC AGG ACC GTG AAG GCG AAG TGA					
K E A L R R T V K G K -					

Translation begun with base no. 75

Translated to base no.1010

Sequence printed from base no. 75 to base no. 1010

Sequence numbered beginning with base no. 75

Figure 12A Translated sequence of F12T.D1S

	10	20	30	40	50	60
ATG GAA TCA GGG AAC AGC ACA AGA AGA TTT TCA AGT TTT TTT CTT CTT GCA TTT ACA GAA	*	*	*	*	*	*
M E S G N S T R R F S S F F L L G F T E						
	70	80	90	100	110	120
AAC CCA CAA CTT CAC TTC CTC ATT TTT GCA CTA TTC CTG TCC ATG TAC CTG GTA ACA CTG	*	*	*	*	*	*
N P Q L L H F L L I F A L F L L S M Y L V T V						
	130	140	150	160	170	180
CTT GGG AAC CTG CTT ATC ATT ATG GCC ATC ATC ACA CAG TCT CAT TTG CAT ACA CCC ATG	*	*	*	*	*	*
L G N L L I I M A I I T Q S H L H T P M						
	190	200	210	220	230	240
TAC TTT TTC CTT GCT AAC CTA TCC TTT GTG GAC ATC TGT TTC ACC TCC ACC ATC CCA	*	*	*	*	*	*
Y F F L A N L S F V D I C F T S T I P						
	250	260	270	280	290	300
	*	*	*	*	*	*

Figure 12B

AAG ATG TTG GTA AAT ATA TAC ACC CAG AGC AAG AGC ATC ACC TAT GAA GAC TGT ATT AGC
K M L V N I Y T Q S K S I T Y E D C I S

310 * 320 * 330 * 340 * 350 * 360 *

CAG ATG TGT CTC TTC TTG GTT TTC GCA GAA TTG GCC AAC TTT CTC CTG CCT GTG ATG GCC
Q M C V F L V F A E L G N F L L A V M A

370 * 380 * 390 * 400 * 410 * 420 *

TAT GAC CGA TAT GTG GCT A-C TGT CAC CCA CTC TGT TAC ACA GTC ATT GTG AAC CAC CCG
Y D R Y V A X C H P L C Y T V I V N H R

430 * 440 * 450 * 460 * 470 * 480 *

CTC TGT ATC CTG CTT CTG CTG TCC TGG GTT ATC AGC ATT TTC CAT GCC TTC ATA CAG
L C I L L L L S W V I S I F H A F I Q

490 * 500 * 510 * 520 * 530 * 540 *

AGC TTA ATT GTG CTA CAG TTG ACC TTC TGT TGT GCA GAT GTG AAA ATC CCT CAC TTC TTC TGT
S L I V L Q L L T F C G D V K I P H F C

550 * 560 * 570 * 580 * 590 * 600 *

GAA CTT AAT CAG CTG TCC CAA CTC ACC TGT TCA GAC AAC TTT CCA AGT CAC CTC ATA ATG
E L N Q L S Q Q L T C S D N F P S H L I M

43/99

Figure 12C

610	*	620	*	630	*	640	*	650	*	660	*
AAT CTT GTA CCT GTT ATG TTG GCA GCC ATT TCC TTC AGT GGC ATC CTT TAC TCT TAT TTC											
N L V P V M L A A I S F S G I L Y S Y F											
670	*	680	*	690	*	700	*	710	*	720	*
AAG ATA GTA TCC TCC ATA CAT TCT ATC TCC ACA CTT CAG GGC AAG TAC AAG GCA TTT TCT											
K I V S S I H S I S T V Q G K Y K A F S											
730	*	740	*	750	*	760	*	770	*	780	*
ACT TGT GCC TCT CAC CTT TCC ATT GTC TCC TTA TTT TAT AGT ACA GGC CTC GGA GTG TAC											
T C A S H L S I V S L F Y S T G L G V Y											
790	*	800	*	810	*	820	*	830	*	840	*
GTC AGT TCT GCT GTG GTC CAA AGC TCA CAT TCT GCT GCA AGT GCT TCG GTC ATG TAT ACT											
PRONUC/TRA	OPTION										
V S S A V V Q S S H S A A S A S V M Y T											

44/99

Figure 12D

850	860	870	880	890	900
* GTG CTC ACC CCC ATG CTG AAC CCC TTC ATT TAT AGT CTA AGG AAT AAA GAT GTG AAG AGA	* V V T P M L N P F I Y S L R N K D V K R	* 910	* 920	* 930	* 940
* GCT CTG GAA AGA CTG TTA GAA GGA AAC TGT AAA CTG CAT CAT TCG ACT GGA TGA	* A L E R L L E G N C K V H H T C -				

Translation begun with base no. 173

Translated to base no.1126

Sequence printed from base no. 173 to base no.1126

Sequence numbered beginning with base no. 173

Figure 13A Translated sequence of I3T.D1S

10 20 30 40 50 60
* * * * *
ATG AAC AAT CAA ACT TTC ATC ACC CAA TTC CTT CTC CTG GGA CTG CCC ATC CCT GAA GAA
M N Q T F I T Q F L L L G L P I P E E
70 80 90 100 110 120
* * * * *
CAT CAG CAC CTG TTC TAT GGC TTG TTC CTG GTC ATG TAC CTC ACC ACC ATC TTG GGA AAC
H Q L F Y A L F L V M Y L T T I L G N
130 140 150 160 170 180
* * * * *
TTG CTA ATC ATT GTA CTT GTT CAA CTG GAC TCC CAG CTC CAC ACA CCT ATG TAT TTG TTT
L L I I V L V Q L D S Q L H T P M Y L F
190 200 210 220 230 240
* * * * *
CTC AGC AAT TTG TCT TTC TCT GAT CTA TGT TTT TCC TCT GTC ACA ATG CCC AAG CTC CTC
L S N L S F S D L L C C F S S V T M P K L L
250 260 270 280 290 300
* * * * *
CAG AAC ATG AGG AGC CAG GAC ACA TCC ATT CCC TAT GGA GGC TGC CTG GCA CAA ACA TAC
Q N M R S Q D T S I P Y G G C L A Q T Y

Figure 13B

310 * 320 * 330 * 340 * 350 * 360 *
TTC TTT ATG GTT TTT GGA GAT ATG GAG AGT TTC CTT CTT GTG GCC ATG GCC TAT GAC CGC
F F M V F F G G D M E S F L L V A M A Y D R

370 * 380 * 390 * 400 * 410 * 420 *
TAT GTG GCC ATO TGC TTC CCT CTG CAT TAC ACC AGC ATC ATG AGC CCC AAG CTC TGT ACT
Y V A I C F P L H Y T S I M S P K L C T

430 * 440 * 450 * 460 * 470 * 480 *
TGT CTA GTG CTG TTA TTG TGG ATG CTG ACG ACA TCC CAT GCC ATG ATG CAC ACA CTG CTT
C L V L L L W M L T T S H A M M H T L L

490 * 500 * 510 * 520 * 530 * 540 *
GCA GCA AGA TTG TCT TTT TGT GAG AAC AAT GTG CTC CTC AAC TTC TTC TGT GAC CTA TTT
A A R L L S F C E N N V V L N F F C D L F

550 * 560 * 570 * 580 * 590 * 600 *
GTT CTC CTA AAG CTG GCC TGC TCA GAC ACT TAT AAT GAG TTG ATG ATA TTT ATC ATG
V L L K L A C S D T Y I N E L M I F I M

610 620 630 640 650 660

Figure 13C

AGT AQA CTC CTC ATT ATT ATT CCA TTC TTC CTC ATT GTT ATG TCC TAT GCA AGG ATC ATA	*	*	*	*	*	*	*
S T L L I I I I P F F F L I I V M S Y A R I I							
670	680	690	700	710	720		
TCC TCT ATT CTT AAG GTT CCA TCT ACC CAA GGC ATC TGC AAG GTC TTC TCT ACC TGT GGT	*	*	*	*	*	*	*
S I L K V P S T Q C I C K V F S T C C							
730	740	750	760	770	780		
TCC CAT CTG TCT GTA GTA TCA CTG TTC TAT GGG ACA ATT ATT GGT CTC TAC TTA TGT CCA	*	*	*	*	*	*	*
S H L S V V S S L F Y G T I I I G L Y L C P							

Figure 14A Translated sequence of I7T.D1S

10 20 30 40 50 60
* * * * *
ATG GAG CGA AGG AAC CAC AGT AGT GGG AGA GTG AGT GAA TTT GTG TTG CTG GGT TTC CCA CCT
M E R R N H S G R V S E F V L L G F P A
70 80 90 100 110 120
* * * * *
CCT GCC CCA CTG CGA GTA CTA CTA TTT TTC CTT TCT CTG G-C TAT GTG TTG GTG TTC
P A P L R V L L F F L S L L L X Y V L V L
130 140 150 160 170 180
* * * * *
ACT GAA AAC ATG CTC ATC ATT ATA GCA ATT AGG AAC CAC CCA ACC CTC CAC AAA CCC ATG
T E N M L I I I A I R N H P T L H K P M
190 200 210 220 230 240
* * * * *
TAT TTT TTC GCT AAT ATG TCA TTT CTG GAG ATT TGG TAT GTC ACT GTT ACC ATT CCT
Y F F L A N M S F L E I W Y V T V T I P
250 260 270 280 290 300
* * * * *
AAG ATG CTC GCT GCC TTC ATT GGT TCC AAG GAG AAC CAT GGA CAG CTG ATC TCC TTT GAG
K M L A G F I G S K E N H G Q L I S F E

Figure 14B

310 * 320 * 330 * 340 * 350 * 360 *
GCA TGC ATG ACA CAA CTC TAC TTT TTC CTG GGC TTG GGT TGC ACA GAG TGT GTC CTT CTT
A C M T Q L Y F L G L G C T E C V L L
370 * 380 * 390 * 400 * 410 * 420 *
GCT GTG ATG GCC TAT GAC CGC TAT GTG GCT ATC TGT CAT CCA CTC CAC TAC CCC GTC ATT
A V M A Y D R Y V A I C H P L H Y P V I
430 * 440 * 450 * 460 * 470 * 480 *
GTC AGT AGC CGG CTA TGT GTG CAG ATG GCA CCT GGA TCC TGG GCT GGA GGT TTT GGT ATC
V S S R L C V Q M A A G S W A G G F G I
490 * 500 * 510 * 520 * 530 * 540 *
TCC ATG GTT AAA GTT TTC CTT ATT TCT CGC CTG TCT TAC TGT GGC CCC AAC ACC ATC AAC
S M V K V F L I S R L S Y C G P N T I N
550 * 560 * 570 * 580 * 590 * 600 *
CAC TTT TTC TGT GAT GTG TCT CCA TTG CTC AAC CTG TCA TGC ACT GAC ATG TCC ACA GCA
H F F C D V S P L L N L S C T D M S T A

50/99

Figure 14C

GAG CTT ACA GAC TTT GTC CTG GCC ATT TTT ATT CTG CTG GGA CCG CTC TCT GTC ACT GCG	610	620	630	640	650	660
E L T D F V L A I F I L L G P L S V T C	*	*	*	*	*	*
GCA TCC TAC ATG CCC ATC ACA GGT GCT GTC ATG CCC ATC CCC TCA GCT GCT GGC CGC CAT	670	680	690	700	710	720
A S Y M A I T G A V M R I P S A G R H	*	*	*	*	*	*
AAA GCC TTT TCA ACC TGT GCC TCC CAC CTC ACT GTT GTG ATC ATC TTC TAT GCA GCC ACT	730	740	750	760	770	780
K A F S T C A S H L T V V I I F Y A A S	*	*	*	*	*	*
ATT TTC ATC TAT GCC AGG CCT AAG GCA CTC TCA GCT TTT GAC ACC AAC AAG CTG GTC TCT	790	800	810	820	830	840
I F I Y A R P K A L S A F D T N K L V S	*	*	*	*	*	*
GTA CTC TAC GCT ATT GTA CCG TTG TTC AAT CCC ATC ATC TAC TGC TTG CGC AAC CAA	850	860	870	880	890	900
PRONUC/TRA OPTION	*	*	*	*	*	*

V L Y A V I V P L F N P I I Y C L R N Q

52/99

Figure 14D

910	*	920	*	930	*	940	*	950	*	960	*								
GAT	GTC	AAA	AGA	CCG	CTA	CGT	CGC	ACG	CTG	CAC	CTG	CCC	CAG	GAC	GAG	CCC	AAT	ACC	
D	V	K	R	A	L	R	R	T	L	H	L	A	Q	D	Q	E	A	N	T
970	*	980	*																
AAC	AAA	GCC	AGC	AAA	ATT	GGT	TAG												
N	K	G	S	K	I	G	-												

Translation begun with base no. 119
Translated to base no.1102
Sequence printed from base no. 119 to base no.1102
Sequence numbered beginning with base no. 119

SUBSTITUTE SHEET

Figure 15A Translated sequence of I8T.D1S

10 * 20 * 30 * 40 * 50 * 60 *
ATG AAC AAC AAA ACT GTC ATC ACC CAT TTC CTC CTC CTG GGA TTG CCC ATC CCC CCA CAG
M N N K T V I T H F L L L L G L P I P P E
70 * 80 * 90 * 100 * 110 * 120 *
CAC CAG CAA CTG TTC TTT GCC CTG TTC CTC ATC ATG TAC CTC ACC ACC TTT CTG GGA AAC
H Q Q L F F A L L F L I M Y L T T F L G N
130 * 140 * 150 * 160 * 170 * 180 *
CTG CTA ATT GTT CTC CTT GTT CAA CTG GAC TCT CAT CTC CAC ACA CCC ATG TAC TTG TTT
L L I V V L V Q L D S H L H T P M Y L F
190 * 200 * 210 * 220 * 230 * 240 *
CTC AGC AAC TTG TCC TTC TCT TCT GAT CTC TGC TGC TTT TCC TCT GTT ACA ATG CTG AAA TTG CTC
L S N L S F S D L L C C F S S V T M L K L L
250 * 260 * 270 * 280 * 290 * 300 *
CAA AAT ATA CAG AGC CAA GTA CCA TCT ATA TCC TAT GCA GGA TGC CTC ACA CAG ATA TTC
Q N I Q S Q V P S I S Y A G C L T Q I F

53/99

SEQUENCE LISTING

[illegible]

Figure 15C

* * * * *
 GGC GTG ATC ATC ATT GTT ATT CCA TTC GTG CTC ATT GTT ATA TCC TAT GCC AAG ATC ATC *
 G V I I I I V I I I V I S Y A K I I
 670 680 690 700 710 720
 * * * * *
 TCC TCC ATT CTT AAG GTT CCA TCT ACT CAA AGC ATT CAC AAG GTC TTC TCC ACT TGT GGT *
 S I L K V P S T Q S I H K V F S T C G
 730 740 750 760 770 780
 * * * * *
 TCT CAT CTC TCT GTG GTG TCT CTG TTC TAC GCG ACA ATT ATT GGT CTC CTC TAT TTA TGT CCA *
 S H L S V V S S L F Y Y C T I I I G L Y L C P
 790 800 810 820 830 840
 * * * * *
 TCA GGT GAT AAT TTT AGT CTA AAG GCG TCT CCC ATG GCT ATG ATG TAC ACA GTG GTA ACT *
 PRONUC/TRA OPTION
 S G D N F S L K G S A M A M M Y T V T
 850 860 870 880 890 900
 * * * * *
 CCA ATG CTG AAC CCG TTC ATC TAC AGC CTA AGA AAC AGA GAC ATG AAG CAG GCC CTA ATA *
 P M L N P F I Y S L R N R D M K Q A L I

Figure 15D

	910		920		930		9
	*		*		*		
AGA GTT ACC TGT AGC AAG AAA ATC TCT CTG CCA TCG TAG							
R V T C S K K I S L P W -							

Translation begun with base no. 57
Translated to base no. 995
Sequence printed from base no. 57 to base no. 995
Sequence numbered beginning with base no. 57

Figure 16A Translated sequence of I9T.D1S

10 20 30 40 50 60
 * * * * *
 ATG ACT AGA AGA AAC CAA ACT GCC ATC TCT CAG TTC TTC CTT CTC GGC CTC CCA TTC CCC
 M T R R N Q T A I S Q F F L L G L P F P

70 80 90 100 110 120
 * * * * *
 GCA GAG TAC CAA CAC CTC TTC TAT GCC CTG TTC CTG GGC ATG TAC CTC ACC ACT CTC CTC
 P E Y Q H L F Y A L F L A M Y L T T L L

130 140 150 160 170 180
 * * * * *
 GGG AAC CTC ATC ATC ATC CTC ATT CTA CTC GAC TCC CAT CTC CAC ACA CCC ATG TAC
 G N L I I I I L L L S H L H T P M Y

190 200 210 220 230 240
 * * * * *
 TTG TTT CTC AGC AAT TTA TCC TCC GAC CTC TGT TTT TCC TCT CTC ACA ATG CCC AAG
 L F L S N L S S F A D L C F S S V T M P K

250 260 270 280 290 300

57/99

Figure 16B

TTG TTG CAG AAC ATG CAG AGC CAA GTT CCA TCC ATC CCC TAT GCA GCG TGC CTG GCA CAG	*	*	*	*	*	*
L L Q N M Q S Q V P S I P Y A G C L A Q						
310	*	320	*	330	*	340
ATA TAC TTC TTT CTG TTT TTT GGA GAC CTT GGA AAC TTC CTG CTT GTG GCC ATG GCC TAT						
I Y F F L F F G D L L G N F L L V A M A Y						
370	*	380	*	390	*	400
GAC CGC TAT GTG GCC ATC TGC TGC CCC CTT CAT TAC ATG AGC ATC ATG AGC CCC AAG CTC						
D R Y V A I C F P L L H Y M S I M S P K L						
430	*	440	*	450	*	460
TGT GTG AGT CTG GTG CTG TCC TGC TGC CTG ACT ACC TTC CAT GCC ATG CTG CAC ACC						
C V S L V V' L S W V V L T T F H A M L H T						
490	*	500	*	510	*	520
CTG CTC ATG GCC AGA TTG TCA TTC TGT GAG GAC AGT GTG ATC CCT CAC TAT TTC TGT GAT						
L L M A R L S F C E D S V I P H Y F C D						
550	*	560	*	570	*	580
ATG TCT ACT CTG CTG AAA CTG CCT TGT TCT GAC ACC CAT GAT AAT GAA TTA GCA ATA TTT						
M S T L L K V A C S D T H D N E L A I F						

58/99

Figure 16C

										59/99											
610	*	620	*	630	*	640	*	650	*	660	*										
ATC TTA GGG GGC CCT ATA GTT GTA CTA CCT TTC CTT CTC ATC ATT GTT TCT TAT GCA AGA	I L G G P I V V L P F L L I I V S Y A R																				
670	*	680	*	690	*	700	*	710	*	720	*										
ATT GTT TCC TCC ATC TTC AAG GTC CCT CCT TCT TCT CAA AGC ATC CAT AAA GCC TTC TCC ACC	I V S S I F K V P S S Q S I H K A F S T																				
730	*	740	*	750	*	760	*	770	*	780	*										
TGT GGC TCC CAC CTG TCT GTG CTG TCA CTG TTC TAT GGG ACA GTC ATT GGT CTC TAC TTA	C G S H L S V V S S L L F Y G T V V I C L Y L																				
790	*	800	*	810	*	820	*	830	*	840	*										
TGT CCT TCA GCT AAT AAC TCC ACT GTG AAG GAG ACT GTC ATG TCT TTG ATG TAC ACA ATG																					
PRONUC/TRA										OPTION											
C	P	S	A	N	N	S	T	V	K	E	T	V	M	S	L	M	Y	T	M		

Figure 17A Translated sequence of I14T.D1S

10 * 20 * 30 * 40 * 50 * 60 *
 ATG ACT GGA AAT AAC CAA ACT TTG ATC TTC GAG TTC CTC CTC GGT CTC CCC ATC CCA
 M T G N N Q T L I L E F L L G L P I P

 70 * 80 * 90 * 100 * 110 * 120 *
 TCA GAG TAT CAT CTC CTG TTC TAT GCC CTG TTC CTG GCC ATG ATG TAC CTC ACC ATC ATC CTG
 S E Y H L L F Y A L F L A M Y L T I I L

 130 * 140 * 150 * 160 * 170 * 180 *
 GGA AAC CTG CTA ATC ATT GTC CTT GTT CGA CTG GAC TCT CAT CTC CTC CAC ATG CCC ATG TAC
 G N L L I I V L R L D S H L H M P M Y

 190 * 200 * 210 * 220 * 230 * 240 *
 TTG TTT CTC AGC AAC TTG TCC TTC TCT GAC CTC TGC TTT TCC TCT CTC ACA ATG CCC AAA
 L F L S N L S F S D L C F S S V T M P K

 250 * 260 * 270 * 280 * 290 * 300 *
 TTG CTT CAG AAC ATG CAG AGC CAA GTA CCA TCT ATA TCC TAT ACA GGC TGC CTG ACA CAG
 L L Q N M Q S Q P S I S Y T G C L T Q

61/99

Figure 17C

ATC TTG GGT GGA CTC ATC ATT ATT ATC CCA TTC CTA TTA ATT GTT ATG TCC TAT GTT AGA *
 I L G G L I I I I P F L L I V M S Y V R *
 670 680 690 700 710 720
 ATT TTC TTC TCC ATT TTG AAG TTT CCA TCT ATT CAG GAC ATC TAC AAG GTA TTC TCA ACC *
 I F S I L K F P S I Q D I Y K V F S T *
 730 740 750 760 770 780
 TGT GGT TCC CAT CTG TCT CTG CTG ACC TTG TTT TAT GGG ACA ATT TTT GGT ATC TAC TTA *
 C G S H L S V V T L L F Y G T I F C I Y L *
 790 800 810 820 830 840
 TGT CCA TCA GGT AAT AAT TCT ACT GTG AAG GAG ATT GCC ATG CCT ATG ATG TAC ACA CTC *
 PRONUC/TRA OPTION
 C P S G N N S T V K E I A M A M Y T V
 850 860 870 880 890 900
 GTG ACT CCC ATG CTG AAT CCC TTC ATC TAC AGC CTG AGG AAC AGA GAC ATG AAA AGG GCC *
 V T P M L N P F I Y S L R N R D M K R A *

63/99

64/99

Figure 17D

CTA	ATA	AGA	GTT	ATC	TGC	ACT	AAG	AAA	ATC	TCT	CTG	TAA	
L	I	R	V	I	C	T	K	K	I	S	L	-	
													9

Translation begun with base no. 64
Translated to base no.1002
Sequence printed from base no. 64 to base no.1002
Sequence numbered beginning with base no. 64

Figure 18A Translated sequence of I15T.D1S

10	20	30	40	50	60
* ATG ACA GAA GAG AAC CAA ACT GTG ATC TCC CAG TTC CTT CTC CTT TTC CTG CCC ATC CCC	* M T E E N Q T V I S Q F L L L F L P I P	* 70	* 80	* 90	* 100
* TCA GAG CAC CAG CTG TTC TAC GCC CTG TTC CTG TCC ATG ATG TAC CTC ACC ACT GTC CTC	* S E H Q H V F Y A L F L S M Y L T T V L	* 110	* 120	* 130	* 140
* GGG AAC CTC ATC ATC ATC CTC ATT CAC CTG GAC TCC CAT CTC CAC ACA CCC ATG TAC	* G N L I I I I L I H L D S H L L H T P M Y	* 150	* 160	* 170	* 180
* TTG TTT CTC AGC AAC TTG TCC TTC TCC TCT GAT CTC TGC TCC TTT TCC TCT GTT ACG ATG CCC AAG	* L F L S N L S F S D L C F S S S V T M P K	* 190	* 200	* 210	* 220
* TTG TTG CAG AAC ATG CAG ACC CAA GTT CCA TCC ATC CCC TTT GCA GGC TGC CTG ACA CAA	* 230	* 240	* 250	* 260	* 270

65/99

Figure 18B

TTA TAC TTT TAC CTG TAT TTT GCA GAC CTT GAG AGC TTC CTG CTT GTG GCC ATG GCC TAT	310	320	330	340	350	360
L Y F Y L Y Y F A D L L E S F L L L V A M A Y	*	*	*	*	*	*
GAC CGC TAT GTG GCC ATC TGC TTC CCC CTT CAT TAC ATG AGC ATC ATG AGC CCC AAG CTC	370	380	390	400	410	420
D R Y V A I C F P L L H Y M S I M S P K L	*	*	*	*	*	*
TGT GTG AGT CTG GTG CTG TCC TCG GTG CTG ACC ACC TTC CAT CCC ATG CTG CAC ACC	430	440	450	460	470	480
C V S L V V L S W V L T T F H A M L H T	*	*	*	*	*	*
CTG CTC ATG GCC AGA TTG TCA TTC TGT GCG GAC AAT ATG ATC CCC CAC TTC TGT GAT	490	500	510	520	530	540
L L M A R L L S F C A D N M I P H F C D	*	*	*	*	*	*
ATA TCT CCT TTA TTG AAA CTG TCC TCG TCT GAC ACC CAT GTT AAT GAG TTG CTG ATA TTT	550	560	570	580	590	600
I S P L L L K L S C S D T H V N E L V I F	*	*	*	*	*	*
	610	620	630	640	650	660

66/99

SUBSTITUTE SUB-

[illegible]

Figure 18D

	910	920	930	940
CTG ATA AGA GTC CTT TGT AAA AAG AAA ATT ACC TTC TGT CTA TGA	*	*	*	*
L I R V L C K K K I T F C L -				

Translation begun with base no. 8
Translated to base no. 952
Sequence printed from base no. 8 to base no. 952
Sequence numbered beginning with base no. 8

69/99

Figure 19A

Translated Sequence of H5.D1S

10	20
ATC TGT TTT GTG TCT ACC ACT GTC CCA	
I C F V S T T V P	
70	80
*	*
GTC ATC ACC TAT GCA GAC TGC ATC ACC	
V I T Y A D C I T	
*	*
GAC AGC TTA CTC CTG ACT GTG ATG GCC	
D S L L L T V M A	
190	200
*	*
CAC TAC ACA GTC ATT ATG AGC TCC TGG	
H Y T V I M S S W	
250	260
*	*
GTG AGC ATC CTA TAT TCT CTG TTA CAA	
V S I L Y S L L Q	

70/99

Figure 19B

30			40				50			60
*			*				*			*
AAG	CAG	CTG	GTG	AAC	ATC	CAG	ACA	CAG	AGC	AGA
K	Q	L	V	N	I	Q	T	Q	S	R
90			100				110			120
*										
CAG	ATG	TGC	TTT	TTT	ATA	CTC	TTT	GTA	GTG	TTG
Q	M	C	F	F	I	L	F	V	V	L
			160				170			180
*			*				*			*
TAT	GAC	CGG	TTT	GTG	GCC	ATC	TGT	CAC	CCC	CTG
Y	D	R	F	V	A	I	C	H	P	L
210			220				230			240
*			*				*			*
CTC	TGT	GGA	CTG	CTG	GTT	CTG	GTG	TCC	TTG	ATC
L	C	G	L	L	V	L	V	S	N	I
270			280				290			300
*			*				*			*
AGC	ATA	ATG	GCA	TTG	CAG	CTG	TCC	TTC	TGT	ACA
S	I	M	A	L	Q	L	S	F	C	T

71/99

Figure 19C

310 *	320 *	330 *
GAA CTG AAA ATC CCT CAA TTT TTC TGT GAA		
E L K I P Q F F C E		
370 *	380 *	390 *
GAC ACT TTT ATT AAT GAC ATG ATG ATG AAT		
D T F I N D M M M N		
430 *	440	450 *
CTC GCT GGA ATA TTT TAC TAC TTT AAG		
L A G I F Y X Y F K		
490 *	500 *	510 *
GCT CAG GGG ATG AAT AAA GCA CTT TCC ACC		
A Q G M N K A L S T		
550	560 *	570 *
TTT TAT TGT ACA GGC GTA GGT GTG TAC CTT		
F Y C T G V G V Y L		
610 *	620 *	630 *
AAT GCT GCA GCC TCG GTG ATG TAC ACT GTG		
N A A A S V M Y T V		

72/99
Figure 19D

340	350	360
★	★	★
CTT AAT CAG GTC ATC CAC CTT GCC TGT TCC		
L N Q V I H L A C S		
400	410	420
★	★	★
TTT ACA AGT GTG CTG CTG GGT GGG GGA TGC		
F T S V L L G G G C		
460	470	480
★	★	★
ATA CTT TGT TGC ATA TGT TCG ATC TCA TCA		
I L C C I C S I S S		
520	530	540
★	★	★
TGT GCA TCT CAC CTC TCA GTT GTC TCC TTA		
C A S H L S V V S L		
580	590	600
★	★	★
AGT TCT GCT GCA ACC CAT AAC TCA CTC TCA		
S S A A T H N S L S		
640		
★		
GTC ACC TCC ATG CTG		
V T S M L		

73/99

```

CATCTGCTTTACTTCTGCTAGCATCCCAAGATGCTAGTGAAATATACAGACCAGAACA
1-----+-----+-----+-----+-----+-----+60
I C F T S A S I P K M L V N I Q T K N K -
GGTGATCACCTATGAAGGCTGCATCTCCTCCAAGTATACTTTTCATACTTGTGGAGTTTG
61-----+-----+-----+-----+-----+-----+120
V I T Y E G C I S Q V Y F S Y S L E F W -
GACAACTTTCTTCGACTGTGATGGCCTATGACCGATATGTGCCATCTGTCACCCATC
121-----+-----+-----+-----+-----+-----+180
T T F F S T V M A Y D R Y V A I C H P S -
TXACTACAGGTCATCATGAACCXXXXXXXXXXXXXXX
181-----+-----+-----+-----+-----+-----+240
? Y T G H E P ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?

```

קריאת א' חזקתו ודעתו ודעתו ודעתו

75/99

Figure 20C

541 TTCTACACTTTTGGTGTGTACCTTAGTCTCTTTTACCCAAACTCACACTCAACTGC
 S T L L G V Y L S S F T Q N S H S T A -
 601 ACGGCACTCTGTATGTACAGTGTGGTCACCCCATGTTG
 R A S V M Y S V V T P M L -
 640

Figure 21A

1 ACCTCCACCACCATCCCAAGATGCTGGTAATATACACACCAGAACATACTATCACC
 T S T T I P K M L V N I H T Q S N T I T - 60
 61 TATGAAGACTGTATTCCAGATGTTGTACTCTTGGTCTTTCGAGAACTGGACAACCTT
 Y E D C I S Q M F V L L V F C E L D N F - 120
 121 CTCCTGGCTGATGGCTATGATCGATATGIGCTATCTGTACCCACTGTATTACACA
 L L A V M A Y D R Y V A I C H P L Y Y T - 180
 181 GTCATTGTGAACCCGACTCTGTATCCCTGCTTCTGCTGCTGGTGTGTCAGCAT
 V I V N H R L C I L L L L L S W V V S I - 240
 241 TTACATGCCCTTCTTACAGAGCTTAATTGTACTACAGTTGACCTTCTGTGAGATGTGAAA
 L H A F L Q S L I V L Q L T F C G D V K - 300

76/99

77/99

Figure 21B

```
ATCCCTCACTTCTGTGAGCTCAATCAGCTGTCCCAACTCACATGTTTCAGACAACCTTT
301 -----+-----+-----+-----+-----+-----+-----+360
      I P H F F C E L N Q L S Q L T C S D N F -
CCAAGTCACCTCACAAATGCATCTTGTACCTGTATATTGCAGCTATTTCCTCAGTCGCT
361 -----+-----+-----+-----+-----+-----+-----+420
      P S H L T M H L V P V I F A A I S L S Q -
ATCCTTTACTCTTATTTCAGATAGTGTCTTCCATACGTTCTATGTCCCTCAGTTCAAGCG
421 -----+-----+-----+-----+-----+-----+-----+480
      I L Y S Y F K I V S S I R S M S S V Q G -
AAGTACAAGGCATTTCCTACATGTGCCCTCTCACCTTTCCTTGTCTCTTATTATTATAGT
481 -----+-----+-----+-----+-----+-----+-----+540
      K Y K A F S T C A S H L S I V S L F Y S -
ACAGGCCCTCGGGGTGACGTACGTTCTGTGATCCGAAGCTCACACTCTCTGCAAGT
541 -----+-----+-----+-----+-----+-----+-----+600
      T G L G V Y V S S A V I R S S H S S A S -
GCTTCGGTCATGTATACGTGGTCACCCCATGTTG
601 -----+-----+-----+-----+-----+-----+-----+636
      A S V M Y T V V T P M L -
```

Figure 22A

CATAGCTATTCATCTTCTGTCACACCCCAATATGCTGTCAACTTCCTTATAAGCAAAA	1	60
I G Y S S V T P N M L V N F L I K Q N -		
TACCATCTCATACCTTGGATGTTCTATACAGTTTGGCTCAGCTGCTTGTGTTGACGCTCT	61	120
T I S Y L G C S I Q P G S A A L F G G L -		
TGAATGCTTCTCTGCTGCCATGGCGTATGATCGTTTGTAGCAATCTGCAACCCACT	121	180
E C F L L A A M A Y D R F V A I C N P L -		
GCTTTATTCAACGAATAATGTCCACACAAGTCTGTGTCAGTTGGTGTGGGATCTTATAT	181	240
L Y S T K M S T Q V C V Q L V V G S Y I -		
AGGGGATTCTTAATGCCCTGCTCTTTTACCCTTTCCTTTTTCCTTGTCTCTGTCGG	241	300
G G F L N A S S P T L S F F S L S P C G -		

79/99

Figure 22B

```
ACCAAATAGAAATCAACTTTTACTGTGATTTTGCTCCGTTAGTAGAACTTTCTTGCTTC
301 -----+-----+-----+-----+-----+-----+360
      P N R I N H F Y C D F A P L V E L S C S -
TGATGTCAGTGTTCCTGATGCTGTACCTCATTTTCTGCTGCTCAGTTACTATGCTCAC
361 -----+-----+-----+-----+-----+-----+420
      D V S V P D A V T S F S A A S V T M L T -
AGTGTATTATAGCCATCTCCTATACCTATATCCTCATCACCATCCTGAAGATCGGTTTC
421 -----+-----+-----+-----+-----+-----+480
      V F I I A I S Y T Y I L I T I L K H R S -
CACTGAGGGTGGACAGAAAGCATTTCTACTGCGACTTCCACCCTCACTGCAGTCACCTCT
481 -----+-----+-----+-----+-----+-----+540
      T E G R Q K A P S T C T S H L T A V T L -
GTGCTATCGAACCATCACATTCATCTATGTGATGCCCAAGTCCAGCTACTCCACAGACCA
541 -----+-----+-----+-----+-----+-----+600
      C Y C T I T F I Y V M P K S S Y S T D Q -
GAACAAGGTGCTCTGCTGTTTATATGGTGGTGATCCCATGTTG
601 -----+-----+-----+-----+-----+-----+646
      N K V V S V F Y M V V I P M L -
```

Figure 23A

1 CATCTGCAAGCCCTGCACTACACCACCATCATCAATAACCGAGTGTGCACAGTTCTAGT 60
I C K P L H Y T T I M N N R V C T V L V -
61 CCTCTCCTGTTGGCTGGCTGGTGTGATCATCTCTCCACCTCTTGGTCAATGCCCTCCA
L S C W F A G L L I I L P P L O H G L Q -
121 GCTCGAGTCTGTGACTCCAATGTGATGATCATTTTCGCTGTGATGCCCTCTCCAATTCT 180
L E P C D S N V I D H F G C D A S P I L -
181 GCAGATAACCTGCTCAGACACCGTATTATAGAGAAATTGCTTGGCTTTTGCCATACT 240
Q I T C S D T V F I B K I V L A F A I L -
241 GACACTCATCATTACTCTGCTGATGTGTGTTCTCTCTACACATACATCAAGACCAT 300

80/99

Figure 23B

301 TTTAAGTTTCCTTCTGCTCAACAAGAAAGGCCCTTTTCTACATGTTCTTCCACAT
 L K P P S A Q Q P K K A F S T C S S H M - +360

361 GATTGTGGTTTCATCACCTATGGAGCTGTATTTCATCTACATCAACCTTCAGCGAA
 I V V S I T Y G S C I F I Y I K P S A K - +420

421 GGAAGGGTAGCCATCAATAAGGTTGTATCTGTGCTCACAACATCAGTCGCCCTTTGCT
 E G V A I N K V V S V L T T S V A P L L - +480

C
 481 - 481
 G

82/99

Figure 24A

```

1  CATCTGCCACCGCTCCACTACTCTCTCTCTCATGAGTCCTGACAACTGTGCTGCTCTGGT
   -----+-----+-----+-----+-----+-----+-----+-----+
   I C H P L H Y S L L M S P D N C A A L V - 60
61  AACAGTCTCTGGGTGACACGGGTGGGCACGGGCTTCTGCTTCCCTTCCCTGATTCTAA
   -----+-----+-----+-----+-----+-----+-----+-----+
   T V S W V T G V G T G F L P S L L I S K - 120
121 GTTGGACTTCTGTGGGCCCAACCGCATCAACCATTTCTTCTGTGACCTTCCCTCCATTAAAT
   -----+-----+-----+-----+-----+-----+-----+-----+
   L D F C G P N R I N H F P C D L P P L I - 180
181 CCAGCTGTCTGCTCCAGCGTCTTTGTGACAGAAATGGCCATCTTTGTCTGTCCATCGC
   -----+-----+-----+-----+-----+-----+-----+-----+
   Q L S C S S V P V T E M A I F V L S I A - 240

```

J8

83/99

Figure 24B

241 TGTGCTCTGCATCTGTTTCCTCCTAACCCXXXTCCTACATTTTCATAGTCTCCTCCAT
-----+-----+-----+-----+-----+-----+300
V L C I C F L L T ? ? S Y I F I V S S I -

301 TCTGAGATCCCTTCCACTACCGCAGGATGAAGACATTTTCTACATGTGGCTCCACCT
-----+-----+-----+-----+-----+-----+360
L R I P S T T G R M K T F S T C G S H L -

361 GGCCGTGTCACCATCTACTATCGGACCATGATCTCCATGTATGTCCGCCAATGCCCA
-----+-----+-----+-----+-----+-----+420
A V V T I Y Y G T M I S M Y V G P N A H -

421 TCTGTCCCGAGCTCAACAAGTCAATTCTGTCTTCTACACTGTGATCACCCTACT
-----+-----+-----+-----+-----+-----+480
L S P E L N K V I S V F Y T V I T P L L -

G
481 - 481

Figure 25A

115

[illegible]

67 / 98

Figure 18C

GTC	ATG	GGA	GGG	CTT	GTT	ATT	GTC	ATT	CCA	TTT	GTG	CTC	ATC	ATT	GTA	TCT	TAT	GCA	CGA	*
V	M	G	G	L	V	I	V	I	P	F	V	L	I	I	V	S	Y	A	R	*
670				680		690		700		710		720								
GTT	GTC	GGC	TCC	ATT	CTT	AAA	GTC	CCT	TCT	GTC	CGA	GGC	ATC	CAC	AAG	ATC	TTC	TCC	ACC	*
V	V	A	S	I	L	K	V	P	S	V	R	G	I	H	K	I	F	S	T	*
730				740		750		760		770		780								
IGC	GGC	TCC	CAT	CTG	TCT	GTG	GTG	TCA	CTG	TTC	TAT	GGG	ACA	ATC	ATT	GGT	CTC	TAC	TTA	*
C	G	S	H	L	S	V	V	S	L	F	Y	G	T	I	I	G	L	Y	L	*
790				800		810		820		830		840								
CGT	CCG	TCA	GCT	AAT	AAG	TCT	ACT	GTG	AAG	GAG	ACT	GTC	ATG	CCC	ATG	ATG	TAC	ACA	GTG	*
PRONUC/TRA				OPTION																
C	P	S	A	N	N	S	T	V	K	E	T	V	M	A	M	M	Y	T	V	
850				860		870		880		890		900								
TG	ACC	CCC	ATG	CTG	AAC	CCC	TTC	ATC	TAC	AGC	CTG	AGG	AAC	AGA	GAC	ATG	AAA	GAG	GCA	*
V	T	P	M	L	N	P	F	I	Y	S	L	R	N	R	D	M	K	E	A	*

FIGURE 19

FIGURE 19A
FIGURE 19B
FIGURE 19C

FIGURE 19A

Translated Sequence of H5.D1S

10			Tm2			20			30		
*						*			*		
ATC	TGT	TTT	GTG	TCT	ACC	ACT	GTC	CCA	AAG	CAG	
I	C	F	V	S	T	T	V	P	K	Q	
40			50			60					
*			*			*					
CTG	GTG	AAC	ATC	CAG	ACA	CAG	AGC	AGA	GTC	ATC	
L	V	N	I	Q	T	Q	S	R	V	I	
70			80			90			Tm3		
*			*			*			*		
ACC	TAT	GCA	GAC	TGC	ATC	ACC	CAG	ATG	TGC	TTT	
T	Y	A	D	C	I	T	Q	M	C	F	
110			120			130					
*			*			*					
TTT	ATA	CTC	TTT	GTA	GTG	TTG	GAC	AGC	TTA	CTC	
F	I	L	F	V	V	L	D	S	L	L	
140			HaeIII			P1			160		
*			*			*			*		
CTG	ACT	GTG	ATG	GCC	TAT	GAC	CGG	TTT	GTG	GCC	
L	T	V	M	A	Y	D	R	F	V	A	
170			P9			180			190		
*			*			*					
ATC	TGT	CAC	CCC	CTG	CAC	TAC	ACA	GTC	ATT	ATG	
I	C	H	P	L	H	Y	T	V	I	M	
200			210			220			Tm4		
*			*			*			*		
AGC	TCC	TGG	CTC	TGT	GGA	CTG	CTG	GTT	CTG	GTG	
S	S	W	L	C	G	L	L	V	L	V	

FIGURE 19B

240			250			260				
*			*			*				
TCC	TTG	ATC	GTG	AGC	ATC	CTA	TAT	TCT	CTG	TTA
S	W	I	V	S	I	L	Y	S	L	L
270			280			290				
*			*			*				
CAA	AGC	ATA	ATG	GCA	TTG	CAG	CTG	TCC	TTC	TGT
Q	S	I	M	A	L	Q	L	S	F	C
300			310			320			330	
*			*			*			*	
ACA	GAA	CTG	AAA	ATC	CCT	CAA	TTT	TTC	TGT	GAA
T	E	L	K	I	P	Q	F	F	C	E
340			350			360				
*			*			*				
CTT	AAT	CAG	GTC	ATC	CAC	CTT	GCC	TGT	TCC	GAC
L	N	Q	V	I	H	L	A	C	S	D
370			380			Tm5			390	
*			*			*			*	
ACT	TTT	ATT	AAT	GAC	ATG	ATG	ATG	AAT	TTT	ACA
T	F	I	N	D	M	M	M	N	F	T
400			410			420			430	
*			*			*			*	
AGT	GTG	CTG	CTG	GGT	GGG	GGA	TGC	CTC	GCT	GGA
S	V	L	L	G	G	G	C	L	A	G
440			450			460				
*			*			*				
ATA	TTT	TAC	T--	TAC	TTT	AAG	ATA	CTT	TGT	TGC
I	F	Y	X	Y	F	K	I	L	C	C
470			480			490				
*			*			*				
ATA	TGT	TCG	ATC	TCA	TCA	GCT	CAG	GGG	ATG	AAT
I	C	S	I	S	S	A	Q	G	M	N

FIGURE 19C

	500		510		Tm6		520				
	*		*				*				
AAA	GCA	CTT	TCC	ACC	TGT	GCA	TCT	CAC	CTC	TCA	
K	A	L	S	T	C	A	S	H	L	S	
530			540			550				560	
*			*			*				*	
GTT	GTC	TCC	TTA	TTT	TAT	TGT	ACA	GGC	GTA	GGT	
V	V	S	L	F	Y	C	T	G	V	G	
		570			580				590		
		*			*				*		
GTG	TAC	CTT	AGT	TCT	GCT	GCA	ACC	CAT	AAC	TCA	
V	Y	L	S	S	A	A	T	H	N	S	
	600			610		Tm7		620			
	*			*				*			
CTC	TCA	AAT	GCT	GCA	GCC	TCG	GTG	ATG	TAC	ACT	
L	S	N	A	A	A	S	V	M	Y	T	
630			640								
*			*								
GTG	GTC	ACC	TCC	ATG	CTG						
V	V	T	S	M	L						

FIGURE 20

FIGURE 20A
FIGURE 20B
FIGURE 20C

FIGURE 20A

1 CATCTGCTTACTTCTGCTAGCATCCCAAGATGCTAGTGAATATACAGACGAAGAACAA 60
I C F T S A S I H K M L V N I Q T K N K -

61 GGTGATCACCTATGAAGGCTGCATCTCCCAAGTATACTTTTCATACTCTTTGGAGTTTG 120
V I T Y E G C I S Q V Y F S Y S L E F W -

121 GACAACTTCTTCGACTGTGATGGCCTATGACCGATATGTGGCCATCTGTCACCCATC 180
T T F S T V M A Y D R Y V A I C H P S -

181 TXACTACAGGTCATGAACCXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 240
? Y T G H H E P ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?

FIGURE 21

FIGURE 21A

FIGURE 21B

FIGURE 21A

Tm2

1	ACCTCCACCACCATCCCAAGATGCTGGTAAATATACACCCAGAGCAATACTATCACC	60
	T S T T I P K M L V N I H T Q S N T I T	-
60	TATGAAGACTGTATTCCAGATGTTTGTA	120
	Y E D C I S Q M F V L L V F G E L D N P	-
121	CTCCTGGCTGTGATGGCCTATGATCGATATGTGGCTATCTGTCA	180
	L L A V M A Y D R Y V A I C H P L Y Y T	-
181	GTCATTGTGAACCA	240
	V I V N H R L C I L L L L L S W V S I	-
241	TTACATGCCTTCTTACAGAGCTTAATTGTACTACAGTTGACCTTCTGTGGAGATGTGAAA	300
	L H A F L Q S L I V L Q L T F C G D V K	-

FIGURE 21B

```
301 ATCCCTCACTTCTGTGAGCTCAATCAGCTGTCCCAACTCACATGTTCAGACAAC TT 360 -  
-----+-----+-----+-----+-----+-----+-----+  
I P H F F C E L N Q L S Q L T C S D N F  
361 CCAAGTCACCTCACAATGCATCTTGACCTGTATATTTGCAGCTATTTCCCTCAGTGGT 420 -  
-----+-----+-----+-----+-----+-----+-----+  
P S H L T M H L V P V I F A A I S L S G  
421 ATCCTTACTCTTATTCAAGATAGTGCTTCCATACGTTCTATGTCCCTCAGTTCAGGG 480 -  
-----+-----+-----+-----+-----+-----+-----+  
I L Y S Y F K I V S S I R S M S S V Q G  
481 AAGTACAAGGCATTCTACATGTGCCCTCTCACCTTTCACATGTCTCCTTATTTATAGT 540 -  
-----+-----+-----+-----+-----+-----+-----+  
K Y K A F S T C A S H L S I V S L F Y S  
541 ACAGGCCCTCGGGGTGACGTCTGTGATCCGAAGCTCACACTCCTCTGCAAGT 600 -  
-----+-----+-----+-----+-----+-----+-----+  
T G L G V Y V S S A V I R S S H S S A S  
601 GCTTCGGTCATGTATCTGTGGTCACCCCATGTTG 636  
-----+-----+-----+-----+-----+-----+-----+  
A S V M Y T V V T P M L -
```

FIGURE 22

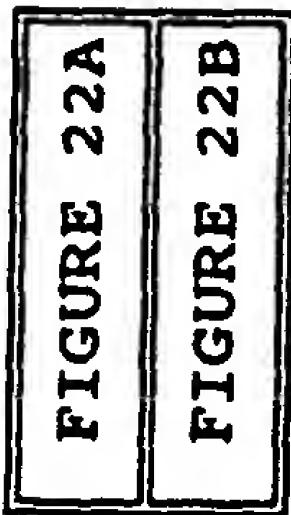


FIGURE 22A

J4

Tm2

```
1  CATAGGCTATTCACTTCTGTCAACCCCAATATGCTTGTCAACTTCCTTATAAGCAAAA 60 -
   I G Y S S S V T P N M L V N F L I K Q N
61  TACCATCTACCTTGGATGTTCTATACAGTTGGCTCAGCTGCTTGTGGAGGTCT 120 -
   T I S Y L G C S I Q F G S A A L F G G L
121 TGAATGCTTCCTTCTGGCTGCCATGGCGTATGATCGTTTGTAGCAATCTGCAACCCACT 180 -
   E C F L L A A M A Y D R F V A I C N P L
181 GCTTTATTCACGAAATGTCCACACAAGTCTGTGTCCAGTTGGTGTGGATCTTATAT 240 -
   L Y S T K M S T Q V C V Q L V V G S Y I
241 AGGGGATTCTTAATGCCCTCCTCTTTTACCCTTTTCCTTTTTCCTTGTCCTTCTGTGG 300 -
   G G F L N A S S F T L S F F S L S F C G
```

FIGURE 22B

```
301 ACCAAATAGAAATCAATCACTTTTACTGTGATTTTGCTCCGTTAGTAGAACTTTCTTGCTC 360 -
    P N R I N H F Y C D F A P L V E L S C S -
361 TGATGTCAGTGTTCCTGATGCTGTACCTCATTTTCTGCTGCCCTCAGTTACTATGCTCAC 420 -
    D V S V P D A V T S F S A A S V T M L T -
421 AGTGTTATCATAGCCATCTCCTATACCTATATCCTCATCACCATCCTGAAGATGCGTTC 480 -
    V F I I A I S Y T Y I L I T I L K M R S -
481 CACTGAGGTCGACAGAAAGCATTTCTACCTGCACCTTCCACCTCACTGCAGTCACTCT 540 -
    T E G R Q K A F S T C T S H L T A V T L -
541 GTGCTATGGAACCATCACATTCTATGTGATGCCCAAGTCCAGCTACTCCACAGACCA 600 -
    C Y G T I T F I Y V M P K S S Y S T D Q -
601 GAACAAGGTGCTGTGTTTATATGTTGGTGATCCCCATGTTG 646 -
    N K V V S V F Y M V V I P M L -
```

FIGURE 23

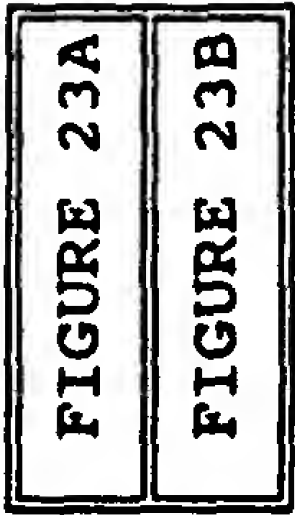


FIGURE 23A

Tm4

```
1  CATCTGCAAGCCCTGCACTACACCACCATCATGAATAACCGAGTGTGCACAGTTCTAGT 60
   I C K P L H Y T I M N N R V C T V L V -
61  CCTCTCCTGTTGGTTGCTGGCCTGTTGATCATCCTCCACCTCTTGGTCATGGCCTCCA 120
   L S C W F A G L L I I L P P L G H G L Q -
121 GCTGGAGTCTGTGACTCCAATGTGATTGATCATTTTGGCTGTGATGCCCTCTCCAATTCT 180
   L E F C D S N V I D H F G C D A S P I L -
181 GCAGATAACCTGCTCAGACACGGTATTATAGAGAAAATTGTCTTGGCTTTTGCCATACT 240
   Q I T C S D T V F I E K I V L A F A I L -
241 GACACTCATTA TCTGGTATGTGTTGTTCTCTCCTACACATACATCAAGACCAT 300
   T L I I T L V C V V L S Y T Y I I K T I
```

FIGURE 23B

301	TTTAAAGTTTCCTTCTGCTCAACAAAGAAAAGGCCCTTTTCTACATGTTCTTCCCACAT	360
	L K F P S A Q Q R K K A F S T C S S H M	-
361	GATTGTGGTTTCCCATCACCTATGGGAGCTGTATTTCATCTACATCAAAACCTTCAGCGAA	420
	I V V S I T Y G S C I F I Y I K P S A K	-
421	GGAAGGGTAGCCATCAATAAGGTTGTATCTGTGCTCACAAACATCAGTCGCCCTTTGCT	480
	E G V A I N K V V S V L T T S V A P L L	-
481	C - 481	
	G	

FIGURE 24

FIGURE 24A

FIGURE 24B

FIGURE 24A

85

	Tm4																					
1	C	A	T	G	C	C	A	C	T	A	C	T	C	T	C	T	C	A	T	G	A	60
	I	C	H	P	L	H	Y	S	L	L	M	S	P	D	N	C	A	A	L	V	-	
61	A	A	C	A	G	T	C	T	G	G	T	G	A	C	A	G	G	G	C	T	T	120
	T	V	S	W	V	T	G	V	G	T	G	F	L	P	S	L	L	I	S	K	-	
121	G	T	T	G	G	A	C	C	G	C	A	T	C	A	A	C	C	A	T	T	180	
	L	D	F	C	G	P	N	R	I	N	H	F	F	C	D	L	P	P	L	I	-	
181	C	C	A	G	T	G	T	C	C	A	G	C	G	T	C	T	T	T	G	T	240	
	Q	L	S	C	S	S	V	F	V	T	E	M	A	I	F	V	L	S	I	A	-	

FIGURE 24B

241 TGTGCTCTGCATCTGTTCCTCCTAACCXXXTCTACATTTTCATAGTGTCCTCCAT 300
-----+-----+-----+-----+-----+-----+-----+
V L C I C F L L T ? ? S Y I F I V S S I -
301 TCTGAGAAATCCCTTCCACTACCGGAGGATGAAGACATTTTCTACATGTGGCTCCACCT 360
-----+-----+-----+-----+-----+-----+-----+
L R I P S T T G R M K T F S T C G S H L -
361 GGCCGTGGTCACCATCTACTATGGGACCATGATCTCCATGTATGTCGGCCCAAATGCGCA 420
-----+-----+-----+-----+-----+-----+-----+
A V V T I Y Y G T M I S M Y V G P N A H -
421 TCTGTCCCGGAGCTCAACAAGGTCATTTCTGTCTTCTACACTGTGATCACCCCACTACT 480
-----+-----+-----+-----+-----+-----+-----+
L S P E L N K V I S V F Y T V I T P L L -
G - 481

FIGURE 25

FIGURE 25A
FIGURE 25B
FIGURE 25C

FIGURE 25A

J11

Tm2

```

2  GTCTGCTTCTCCTCCACCACCTGTCCCAAGGTACTGGCTAACCATACTCAGTAGCA 60
   V C F S S T T V P K V L A N H I L S S Q -
61  GGCCATTTCCTTCTCTGGGTCTCTAACTCAGCTGTATTTCTCTGTGTCTGTGAATAT 120
   A I S F S G C L T Q L Y F L C V S V N M -
121 GGACAAATTTCCTGCTGGCTGTGATGGCCCTATGACAGATTGTGGCCATATGCCACCCTTT 180
   D N F L L A V M A Y D R F V A I C H P L -
181 GTACTACACAAGATGACCCACCAGCTCTGTGTCTGTGGTCTGGATCAXXXXXX 240
   Y Y T T K M T H Q L C V L L V S G S ? -
241 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 300
   ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? -

```

FIGURE 25B

[illegible]

FIGURE 27

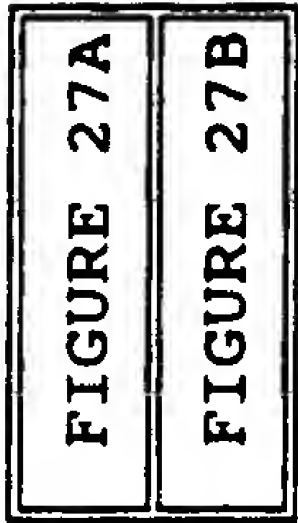


FIGURE 27A

April 5, 1992 00:00 Tm4

J15

1	TATCTGCAACCTCTGCGCTACCCAGTGCTCATGAGCGCGCGGTGTGCTGCTCATGGT	60
	I C N P L R Y P V L M S G R W C L L M V	-
61	CGTGGCCTCCTGGTTGGGAGGATCCCTCAACGCCCTCCATTTCAGACTTCTCTGACCCCTTCA	120
	V A S W L G G S L N A S I Q T S L T L Q	-
121	GTTCCCTACTGTGGATCACGGAAGATCTCCCACTTCTTCTGTGAGGTGCCCTCGCTGCT	180
	F P Y C G S R K I S H F F C E V P S L L	-
181	GAXXTGGCCTGTGCAGACACTGAAGCCTATGAGCAGGTACTATTGTGACAGGCGTGGT	240
	? ? A C A D T E A Y E Q V L F V T G V V	-

FIGURE 27B

```
241      GGTCCCTCCTGGTGCCCATTAATGCTGCTTATGCCCTCATCCTGGCTGCTGT  
-----+-----+-----+-----+-----+-----+-----+  
      V L L V P I T F I T A S Y A L I L A A V      300 -  
  
301      GCTCCGAATGCACTCTGCGGAGGGAGTCAGAAGGCCCTAGCCACATGCTCCTCTCACCT  
-----+-----+-----+-----+-----+-----+-----+  
      L R M H S A E G S Q K A L A T C S S H L      360 -  
  
361      GACAGTCGTCAATCTTCTATGGGCCCCCTTGCTCTACACCTACATGTTACCTGCTTCCTA  
-----+-----+-----+-----+-----+-----+-----+  
      T V V N L F Y G P L V Y T Y M L P A S Y      420 -  
  
421      TCACTCACCAGGCCAAGACGACATAGTATCCGCTCTTTACACCGTTCTCACACCCATGCT  
-----+-----+-----+-----+-----+-----+-----+  
      H S P G Q D D I V S V F Y T V L T P M L      480 -  
  
      T  
481 - 481  
      A
```

FIGURE 28

FIGURE 28A
FIGURE 28B

FIGURE 28A

J16

Tm4

```
1  CATCTGAGGCCCTTCACTATCCTACCTCATGCCAGACACTGTGTGCCAAGATTGC 60 -
   I C R P L H Y P T L M T Q T L C A K I A
61  CACTGGTTGGCTGGAGGCTTGGCTGGGCCAGTGGTAGAAATTTCCTTGGTGCTCG 120 -
   T G C W L G G L A G P V V E I S L V S R
   121  TCTCCTTTTGTGGCCCAATCACATTCACACATCTTTTGTGATTTCCACCTGTGCT 180 -
   L L F C G P N H I Q H I F C D F P P V L
   181  GAGCTTGGCTTGACTGATACATCAGTGAATGTCCTGGTAGATTTTATTATAAACCTCTG 240 -
   S L A C T D T S V N V L V D F I I N L C
   241  CAAGATCCTGGCCACCTTCCTGCTGATCCTGAGCTCCTACTTGAGATAATCCGCACAGT 300 -
   K I L A T F L L I L S S Y L Q I I R T V
```


FIGURE 28B

301 GCTCAAGATTCCCTTCAGCTGCAGGCAAGAAAGCATTCCTCGACTTGTGCTCCCATCT 360
-----+-----+-----+-----+-----+-----+-----+
L K I P S A A G K K A F S T C A S H L -
361 CACTGTGGTTCTCATCTTCTATGGGAGCATCCTTTTCATGTATGTGCGGCTGAAGAAGAC 420
-----+-----+-----+-----+-----+-----+-----+
T V V L I F Y G S I L F M Y V R L K K S -
421 TTACTCCCTTGACTACGACAGAGCCCTTGGCAGTAGTCTACTCCGTGGTTACCCCTTTCCT 420
-----+-----+-----+-----+-----+-----+-----+
Y S L D Y D R A L A V V Y S V V T P F L -
G - 481
481 - 481

FIGURE 29

FIGURE 29A
FIGURE 29B

FIGURE 29A

J17

Tm4

```

1  AATCTGCAACCCACTGCTTTATTCCACCAAAATGTCCACACAAAGTCTGTATCCAGTTGGT 60
   I C N P L L Y S T K M S T Q V C I Q L V
61  TGCAGGATCTTATATAGGGGTTTCTTAATACTTGCCCTCATCATGTTTACTTTTCTC 120
   A G S Y I G G F L N T C L I M F Y F S
121 TTTTCTTCTGTGGCCAAATAGTTGATCATTTTCTGTGATTTTGCTCCTTXXT 180
   F L F C G P N I V D H F F C D F A P ?
181 GGAAC TTGCTGCTGATGTGAGTGTCTGTAGTTGTATGTCA TTTCTGTGGCTC 240
   E L S C S D V S V V V M S F S A G S
241 AGTACTATGATCACAGTGT TATCATAGCCATCTCCTAT TCTTACATCCTCATCACCAT 300
   V T M I T V F I I A I S Y I L I T I

```

FIGURE 29B

301	CCTGAAGATGTCCTCAACTGAGGGCCGTCACAAGGCTTCTCCACATGTACCTCCACCT	360
	L K M S S T E G R H K A F S T C T S H L	-
361	CACTGCAGTCACTCTCTACTATGGCACCATTACCTTCATTATGTGATGCCCAAGTCCAC	420
	T A V T L Y Y G T I T F I Y V M P K S T	-
421	ATACTCTACAGACCAGAACAGGTGGTGTCTGTGTGTTTACATGGTGTGATCCCAATGTT	480
	Y S T D Q N K V V S V F Y M V V I P M L	-
481	G - 481	

FIGURE 30

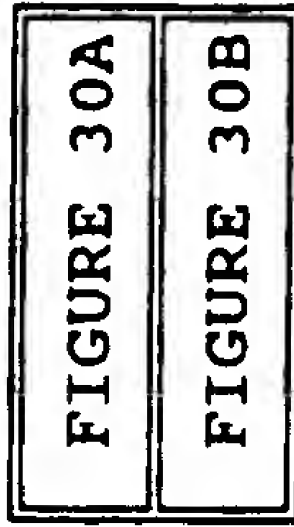


FIGURE 30A

J19

Tm4

1	TATCTGCCACCTCTGAAGTACACAGTTATCATGAATCACTATTTTGTGTGATGCTGCT	60
	I C H P L K Y T V I M N H Y F C V M L L	-
61	GCTCTCTCTGTGTTAGCATTCACATGCGTTGTTCCACATTTAATGGTGTGAT	120
	L F S V F V S I A H A L F H I L M V L I	-
121	ACTGACTTTCAGCACAAAAGTAAATCCCTCACTTTTCTGTGAGCTGGCTCATATCAT	180
	L T F S T K T E I P H F C E L A H I I	-
181	CAAACTTACCTGTTCCGATAATTTATCAACTATCTGCTGATATACACAGTCTGCTT	240
	K L T C S D N F I N Y L L I Y T E S V L	-
241	ATTTTGGTGTTCATATTGTAGGATCATTTTGTCTTATATTACACTGTATCCTCAGT	300
	F F G V H I V G I I L S Y I Y T V S S V	-

FIGURE 31

FIGURE 31A

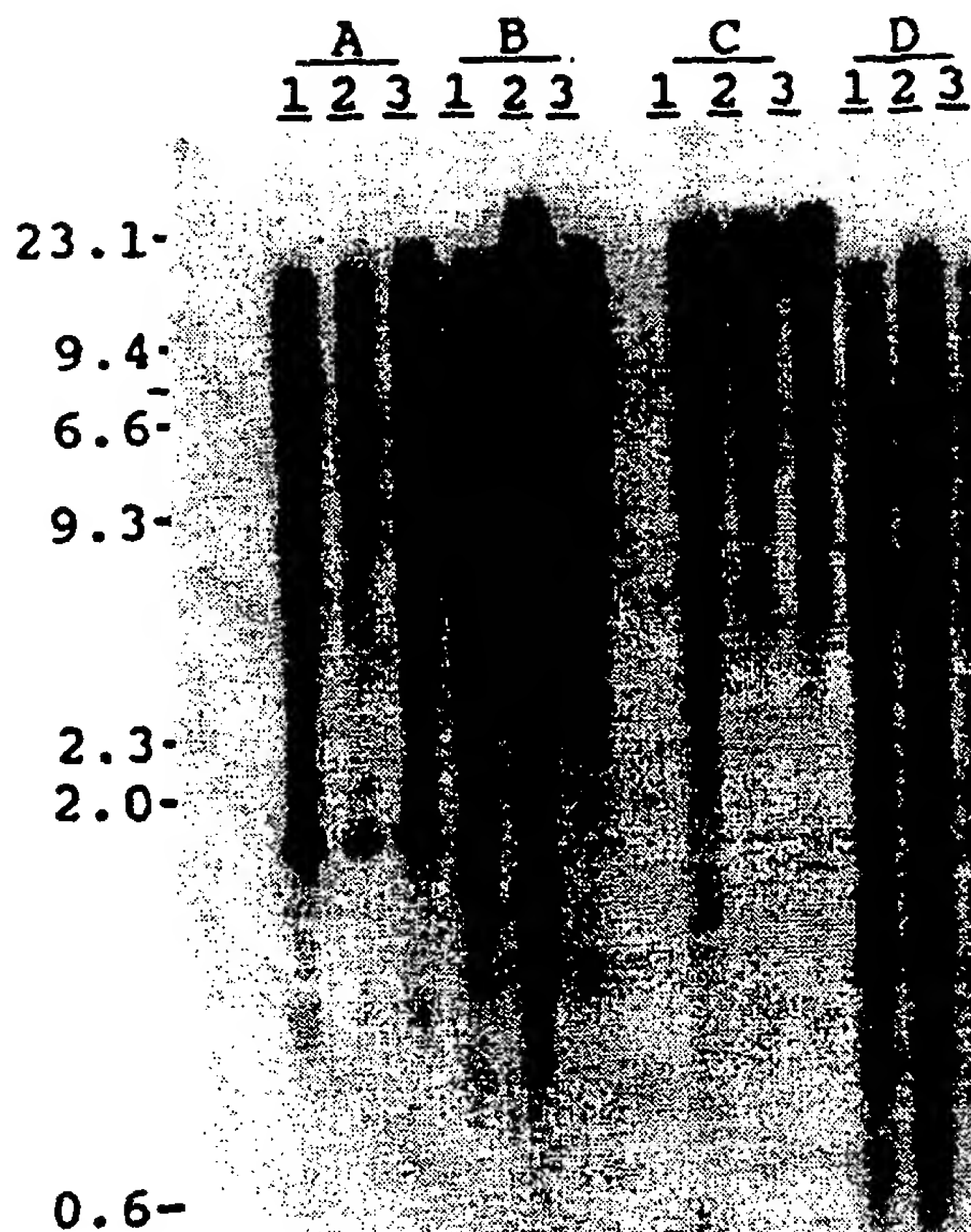
FIGURE 31B

FIGURE 31A

Tm4

1	AATCTGCTACCCACTGAGGTACCTTCTCATCATGAGCTGGGTGGTGTGCACAGCACTGTC	60
	I C Y P L R Y L L I M S W V V C T A L S	
61	CGTGGCAATCTGGGTCATAGGCTTTTGTGCCCTCCGTTATACCTCTCTGCTTCACGATCCT	120
	V A I W V I G F C A S V I P L C F T I L	
121	CCCACTCTGTGGTCCTTACGTCGTTGATTATCTTTTCTGCGAGCTGCCATCCTTCTGCA	180
	P L C G P Y V V D Y L F C E L P I L L H	

Figur 32



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US92/02741

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : C12N 15/12, 15/63, 15/64, 5/10; C07K 13/00; A01N 33/00; A61K 37/00

US CL : 536/27; 424/418; 435/7.21, 172.3, 240.1, 320.1; 514/2; 530/395

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/27; 424/418; 435/7.21, 172.3, 240.1, 320.1; 514/2; 530/395

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CAS ONLINE, MEDLINE, UEMBL, GENBANK, PIR, SWISS PROT, APS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P Y,P	Molecular Brain Research, Volume 13, No. 1-2, issued March 1992, L. A. Selbie et al., "Novel G protein-coupled receptors: a gene family of putative human olfactory receptor sequences," abstract.	1-32 33-98
Y X	Sensory Syst., Volume 1, No. 1, issued 1987, V. I. Novoselov et al., "The properties of receptor molecules from rat olfactory epithelium," abstract.	1-34, 65-98 35-64
X,P Y,P	Nature, Volume 355, issued 30 January 1992, M. Parmentier et al., "Expression of members of the putative olfactory receptor gene family in mammalian germ cells," pages 453-455, see entire document.	1-32 33-98
Y X	Biochimica Biophysica Acta, Volume 839, No. 3, issued 1985, E. E. Fesenko et al., "Molecular mechanisms of olfactory reception. VI Kinetic characteristics of camphor interaction with binding sites of rat olfactory epithelium," abstract.	1-34, 65-98 35-64

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

- * Special categories of cited documents:
- *A* document defining the general state of the art which is not considered to be part of particular relevance
 - *E* earlier document published on or after the international filing date
 - *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 - *O* document referring to an oral disclosure, use, exhibition or other means
 - *P* document published prior to the international filing date but later than the priority date claimed

Date of the actual completion of the international search

25 June 1992

Date of mailing of the international search report

23 July 1992

Name and mailing address of the ISA/
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. NOT APPLICABLE

Authorized officer

LISA T. BENNETT

Telephone No. (703) 308-3988

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/02741

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P Y,P	Chemtracts: Organic Chemistry, Volume 4, No. 4, issued 1991, K. Touhara et al., "A novel multigene family may encode odorant receptors: a molecular basis for odor recognition," abstract.	1-32 33-98
Y,P	Chemical Senses, Volume 16, No. 5, issued 1991, R. H. R. Anholt, "Odor recognition and olfactory transduction: the new frontier," abstract.	1-98
Y	Trends in Neuroscience, Volume 14, No. 7, issued 1991, S. Firestein, "A noseful of odor receptors," abstract.	1-98
Y	Proceedings of the National Academy of Sciences, Volume 86, issued November 1989, E. Dancigier et al., "Olfactory marker protein gene: Its structure and olfactory neuron-specific expression in transgenic mice," pages 8565-8569, see entire document.	1-34
Y	Kagaky Kogyo, Volume 40, No. 11, issued 1989, M. Kashiwayanagi et al., "High sensitivity odor sensor using artificial membrane," abstract.	1-98

-71-

1	5	10	15
Ala Ala Leu Val Thr Val Ser Trp Val Thr Gly Val Gly Thr Gly Phe	20	25	30
Leu Pro Ser Leu Leu Ile Ser Lys Leu Asp Phe Cys Gly Pro Asn Arg	35	40	45
Ile Asn His Phe Phe Cys Asp Leu Pro Pro Leu Ile Gln Leu Ser Cys	50	55	60
Ser Ser Val Phe Val Thr Glu Met Ala Ile Phe Val Leu Ser Ile Ala	65	70	75
Val Leu Cys Ile Cys Phe Leu Leu Thr Xaa Xaa Ser Tyr Ile Phe Ile	85	90	95
Val Ser Ser Ile Leu Arg Ile Pro Ser Thr Thr Gly Arg Met Lys Thr	100	105	110
Phe Ser Thr Cys Gly Ser His Leu Ala Val Val Thr Ile Tyr Tyr Gly	115	120	125
Thr Met Ile Ser Met Tyr Val Gly Pro Asn Ala His Leu Ser Pro Glu	130	135	140
Leu Asn Lys Val Ile Ser Val Phe Tyr Thr Val Ile Thr Pro Leu Leu	145	150	155
			160

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 646 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: J11

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..646

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

N	GTC	TGC	TTC	TCC	TCC	ACC	ACT	GTC	CCC	AAG	GTA	CTG	GCT	AAC	CAC
Val	Cys	Phe	Ser	Ser	Thr	Thr	Val	Pro	Lys	Val	Leu	Ala	Asn	His	
1				5				10					15		

46

SUBSTITUTE SHEET

-72-

ATA CTC AGT AGT CAG GCC ATT TCC TTC TCT GGG TGT CTA ACT CAG CTG Ile Leu Ser Ser Gln Ala Ile Ser Phe Ser Gly Cys Leu Thr Gln Leu 20 25 30	94
TAT TTT CTC TGT GTG TCT GTG AAT ATG GAC AAT TTC CTG CTG GCT GTG Tyr Phe Leu Cys Val Ser Val Asn Met Asp Asn Phe Leu Leu Ala Val 35 40 45	142
ATG GCC TAT GAC AGA TTT GTG GCC ATA TGC CAC CCT TTG TAC TAC ACA Met Ala Tyr Asp Arg Phe Val Ala Ile Cys His Pro Leu Tyr Tyr Thr 50 55 60	190
ACA AAG ATG ACC CAC CAG CTC TGT GTC TTG CTG GTG TCT GGA TCA NNN Thr Lys Met Thr His Gln Leu Cys Val Leu Leu Val Ser Gly Ser Xaa 65 70 75	238
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 80 85 90 95	286
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 100 105 110	334
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 115 120 125	382
NNN NNN NNN NNN NNN NNN NNT GTG ATC ATG GTC ACC CCA TTT GTC TGC Xaa Xaa Xaa Xaa Xaa Xaa Xaa Val Ile Met Val Thr Pro Phe Val Cys 130 135 140	430
ATC CTC ATC TCT TAC ATC TAC ATC ACC AAT GCA GTC CTC AGA GTC TCA Ile Leu Ile Ser Tyr Ile Tyr Ile Thr Asn Ala Val Leu Arg Val Ser 145 150 155	478
TCC TTT AGG GGA GGA TGG AAA GCC TTC TCC ACC TGT GGC TCA CAC CTG Ser Phe Arg Gly Gly Trp Lys Ala Phe Ser Thr Cys Gly Ser His Leu 160 165 170 175	526
GCT GTG GTC TGC CTC TTC TAT GGC ACC ATC ATT GCT GTG TAT TTC AAT Ala Val Val Cys Leu Phe Tyr Gly Thr Ile Ile Ala Val Tyr Phe Asn 180 185 190	574
CCT GTA TCT TCC CAT TCA TCT GAG AAG GAC ACT GCA GCA ACT GTG CTA Pro Val Ser Ser His Ser Ser Glu Lys Asp Thr Ala Ala Thr Val Leu 195 200 205	622
TAC ACA GTG GTG ACT CCC ATG TTG Tyr Thr Val Val Thr Pro Met Leu 210 215	646

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 215 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

SUBSTITUTE SHEET

-73-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Val Cys Phe Ser Ser Thr Thr Val Pro Lys Val Leu Ala Asn His Ile
 1 5 10 15
 Leu Ser Ser Gln Ala Ile Ser Phe Ser Gly Cys Leu Thr Gln Leu Tyr
 20 25 30
 Phe Leu Cys Val Ser Val Asn Met Asp Asn Phe Leu Leu Ala Val Met
 35 40 45
 Ala Tyr Asp Arg Phe Val Ala Ile Cys His Pro Leu Tyr Tyr Thr Thr
 50 55 60
 Lys Met Thr His Gln Leu Cys Val Leu Leu Val Ser Gly Ser Xaa Xaa
 65 70 75 80
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 85 90 95
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 100 105 110
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 115 120 125
 Xaa Xaa Xaa Xaa Xaa Xaa Val Ile Met Val Thr Pro Phe Val Cys Ile
 130 135 140
 Leu Ile Ser Tyr Ile Tyr Ile Thr Asn Ala Val Leu Arg Val Ser Ser
 145 150 155 160
 Phe Arg Gly Gly Trp Lys Ala Phe Ser Thr Cys Gly Ser His Leu Ala
 165 170 175
 Val Val Cys Leu Phe Tyr Gly Thr Ile Ile Ala Val Tyr Phe Asn Pro
 180 185 190
 Val Ser Ser His Ser Ser Glu Lys Asp Thr Ala Ala Thr Val Leu Tyr
 195 200 205
 Thr Val Val Thr Pro Met Leu
 210 215

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 646 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

SUBSTITUTE SHEET

-74-

(vii) IMMEDIATE SOURCE:
(B) CLONE: J14

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..646

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

T GTC TGC TTC TCC TCC ACC ACT GTC CCC AAG GTA CTG GCT AAC CAC	46
Val Cys Phe Ser Ser Thr Thr Val Pro Lys Val Leu Ala Asn His	
1 5 10 15	
ATA CTC AGT AGT CAG GCC ATT TCC TTC TCT GCG TGT CTA ACT CAG CTG	94
Ile Leu Ser Ser Gln Ala Ile Ser Phe Ser Gly Cys Leu Thr Gln Leu	
20 25 30	
TAT TTT CTC TGT GTG TCT GTG AAT ATG GAC AAT TTC CTG CTG GCT GTG	142
Tyr Phe Leu Cys Val Ser Val Asn Met Asp Asn Phe Leu Leu Ala Val	
35 40 45	
ATG GCC TAT GAC AGA TTT GTG GCC ATA TGC CAC CCT TTG TAC TAC ACA	190
Met Ala Tyr Asp Arg Phe Val Ala Ile Cys His Pro Leu Tyr Tyr Thr	
50 55 60	
ACA CCG ATG ACC CAC CAG CTC TGT GTC TTG CTG GTG TCT GGA TCA NNN	238
Thr Pro Met Thr His Gln Leu Cys Val Leu Leu Val Ser Gly Ser Xaa	
65 70 75	
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN	286
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
80 85 90 95	
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN	334
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
100 105 110	
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN	382
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
115 120 125	
NNN NNN NNN NNN NNN NNN NNT GTG ATC ATG GTC ACC CCA TTT GTC TGC	430
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Val Ile Met Val Thr Pro Phe Val Cys	
130 135 140	
ATC CTC ATC TCT TAC ATC TAC ATC ACC AAT GCA GTC CTC AGA GTC TCA	478
Ile Leu Ile Ser Tyr Ile Tyr Ile Thr Asn Ala Val Leu Arg Val Ser	
145 150 155	
TCC TTT AGG GGA GGA TGG AAA GCC TTC TCC ACC TGT GGC TCA CAC CTG	526
Ser Phe Arg Gly Gly Trp Lys Ala Phe Ser Thr Cys Gly Ser His Leu	
160 165 170 175	
GCT GTG GTC TGC CTC TTC TAT GGC ACC ATC ATT GCT GTG TAT TTC AAT	574
Ala Val Val Cys Leu Phe Tyr Gly Thr Ile Ile Ala Val Tyr Phe Asn	
180 185 190	
CCT GTA TCT TCC CAT TCA TCT GAG AAG GAC ACT GCA GCA ACT GTG CTA	622
Pro Val Ser Ser His Ser Ser Glu Lys Asp Thr Ala Ala Thr Val Leu	
195 200 205	

SUBSTITUTE SHEET

-75-

TAC ACA GTG GTG ACT CCC ATG TTG
 Tyr Thr Val Val Thr Pro Met Leu
 210 215

646

(2) INFORMATION FOR SEQ ID NO:26:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 215 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Val Cys Phe Ser Ser Thr Thr Val Pro Lys Val Leu Ala Asn His Ile
 1 5 10 15
 Leu Ser Ser Gln Ala Ile Ser Phe Ser Gly Cys Leu Thr Gln Leu Tyr
 20 25 30
 Phe Leu Cys Val Ser Val Asn Met Asp Asn Phe Leu Leu Ala Val Met
 35 40 45
 Ala Tyr Asp Arg Phe Val Ala Ile Cys His Pro Leu Tyr Tyr Thr Thr
 50 55 60
 Pro Met Thr His Gln Leu Cys Val Leu Leu Val Ser Gly Ser Xaa Xaa
 65 70 75 80
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 85 90 95
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 100 105 110
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 115 120 125
 Xaa Xaa Xaa Xaa Xaa Xaa Val Ile Met Val Thr Pro Phe Val Cys Ile
 130 135 140
 Leu Ile Ser Tyr Ile Tyr Ile Thr Asn Ala Val Leu Arg Val Ser Ser
 145 150 155 160
 Phe Arg Gly Gly Trp Lys Ala Phe Ser Thr Cys Gly Ser His Leu Ala
 165 170 175
 Val Val Cys Leu Phe Tyr Gly Thr Ile Ile Ala Val Tyr Phe Asn Pro
 180 185 190
 Val Ser Ser His Ser Ser Glu Lys Asp Thr Ala Ala Thr Val Leu Tyr
 195 200 205
 Thr Val Val Thr Pro Met Leu
 210 215

(2) INFORMATION FOR SEQ ID NO:27:

- (1) SEQUENCE CHARACTERISTICS:

SUBSTITUTE SHEET

-76-

(A) LENGTH: 481 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: rat olfactory epithelium
 (B) STRAIN: Sprague-Dawley rat
 (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

(B) CLONE: J15

(ix) FEATURE:

(A) NAME/KEY: CDS
 (B) LOCATION: 2..481

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

T ATC TGC AAC CCT CTG CGC TAC CCA GTG CTC ATG AGC GGC CGG GTG	46
Ile Cys Asn Pro Leu Arg Tyr Pro Val Leu Met Ser Gly Arg Val	
1 5 10 15	
TGC CTG CTC ATG GTC GTG GCC TCC TGG TTG GGA GGA TCC CTC AAC GCC	94
Cys Leu Leu Met Val Val Ala Ser Trp Leu Gly Gly Ser Leu Asn Ala	
20 25 30	
TCC ATT CAG ACT TCT CTG ACC CTT CAG TTC CCC TAC TGT GGA TCA CGG	142
Ser Ile Gln Thr Ser Leu Thr Leu Gln Phe Pro Tyr Cys Gly Ser Arg	
35 40 45	
AAG ATC TCC CAC TTC TTC TGT GAG GTG CCC TCG CTG CTG ANN NTG GCC	190
Lys Ile Ser His Phe Phe Cys Glu Val Pro Ser Leu Leu Xaa Xaa Ala	
50 55 60	
TGT GCA GAC ACT GAA GCC TAT GAG CAG GTA CTA TTT GTG ACA GGC GTG	238
Cys Ala Asp Thr Glu Ala Tyr Glu Gln Val Leu Phe Val Thr Gly Val	
65 70 75	
GTG GTC CTC CTG GTG CCC ATT ACA TTC ATT ACT GCC TCT TAT GCC CTC	286
Val Val Leu Leu Val Pro Ile Thr Phe Ile Thr Ala Ser Tyr Ala Leu	
80 85 90 95	
ATC CTG GCT GCT GTG CTC CGA ATG CAC TCT GCG GAG GGG AGT CAG AAG	334
Ile Leu Ala Ala Val Leu Arg Met His Ser Ala Glu Gly Ser Gln Lys	
100 105 110	
GCC CTA GCC ACA TGC TCC TCT CAC CTG ACA GTC GTC AAT CTC TTC TAT	382
Ala Leu Ala Thr Cys Ser Ser His Leu Thr Val Val Asn Leu Phe Tyr	
115 120 125	
GGG CCC CTT GTG TAC ACC TAC ATG TTA CCT GCT TCC TAT CAC TCA CCA	430
Gly Pro Leu Val Tyr Thr Tyr Met Leu Pro Ala Ser Tyr His Ser Pro	
130 135 140	

SUBSTITUTE SHEET

-77-

GGC CAA GAC GAC ATA GTA TCC GTC TTT TAC ACC GTT CTC ACA CCC ATG 478
 Gly Gln Asp Asp Ile Val Ser Val Phe Tyr Thr Val Leu Thr Pro Met
 145 150 155

CTT 481
 Leu
 160

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 160 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Ile Cys Asn Pro Leu Arg Tyr Pro Val Leu Met Ser Gly Arg Val Cys
 1 5 10 15
 Leu Leu Met Val Val Ala Ser Trp Leu Gly Gly Ser Leu Asn Ala Ser
 20 25 30
 Ile Gln Thr Ser Leu Thr Leu Gln Phe Pro Tyr Cys Gly Ser Arg Lys
 35 40 45
 Ile Ser His Phe Phe Cys Glu Val Pro Ser Leu Leu Xaa Xaa Ala Cys
 50 55 60
 Ala Asp Thr Glu Ala Tyr Glu Gln Val Leu Phe Val Thr Gly Val Val
 65 70 75 80
 Val Leu Leu Val Pro Ile Thr Phe Ile Thr Ala Ser Tyr Ala Leu Ile
 85 90 95
 Leu Ala Ala Val Leu Arg Met His Ser Ala Glu Gly Ser Gln Lys Ala
 100 105 110
 Leu Ala Thr Cys Ser Ser His Leu Thr Val Val Asn Leu Phe Tyr Gly
 115 120 125
 Pro Leu Val Tyr Thr Tyr Met Leu Pro Ala Ser Tyr His Ser Pro Gly
 130 135 140
 Gln Asp Asp Ile Val Ser Val Phe Tyr Thr Val Leu Thr Pro Met Leu
 145 150 155 160

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 481 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

SUBSTITUTE SHEET

-78-

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: rat olfactory epithelium

(B) STRAIN: Sprague-Dawley rat

(F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

(B) CLONE: J16

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 2..481

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

C ATC TGT AGG CCT CTT CAC TAT CCT ACC CTC ATG ACC CAG ACA CTG	46
Ile Cys Arg Pro Leu His Tyr Pro Thr Leu Met Thr Gln Thr Leu	
1 5 10 15	
TGT GCC AAG ATT GCC ACT GGT TGC TGG TTG GGA GGC TTG GCT GCG CCA	94
Cys Ala Lys Ile Ala Thr Gly Cys Trp Leu Gly Gly Leu Ala Gly Pro	
20 25 30	
GTG GTA GAA ATT TCC TTG GTG TCT CGT CTC CTT TTT TGT GGC CCC AAT	142
Val Val Glu Ile Ser Leu Val Ser Arg Leu Leu Phe Cys Gly Pro Asn	
35 40 45	
CAC ATT CAA CAC ATC TTT TGT GAT TTC CCA CCT GTG CTG AGC TTG GCT	190
His Ile Gln His Ile Phe Cys Asp Phe Pro Pro Val Leu Ser Leu Ala	
50 55 60	
TGT ACT GAT ACA TCA GTG AAT GTC CTG GTA GAT TTT ATT ATA AAC CTC	238
Cys Thr Asp Thr Ser Val Asn Val Leu Val Asp Phe Ile Ile Asn Leu	
65 70 75	
TGC AAG ATC CTG GCC ACC TTC CTG CTG ATC CTG AGC TCC TAC TTG CAG	286
Cys Lys Ile Leu Ala Thr Phe Leu Leu Ile Leu Ser Ser Tyr Leu Gln	
80 85 90 95	
ATA ATC CGC ACA GTG CTC AAG ATT CCT TCA GCT GCA GGC AAG AAG AAA	334
Ile Ile Arg Thr Val Leu Lys Ile Pro Ser Ala Ala Gly Lys Lys Lys	
100 105 110	
GCA TTC TCG ACT TGT GCC TCC CAT CTC ACT GTG GTT CTC ATC TTC TAT	382
Ala Phe Ser Thr Cys Ala Ser His Leu Thr Val Val Leu Ile Phe Tyr	
115 120 125	
GGG AGC ATC CTT TTC ATG TAT GTG CGC CTG AAG AAG ACT TAC TCC CTT	430
Gly Ser Ile Leu Phe Met Tyr Val Arg Leu Lys Lys Thr Tyr Ser Leu	
130 135 140	
GAC TAC GAC AGA GCC TTG GCA GTA GTC TAC TCC GTG GTT ACC CCT TTC	478
Asp Tyr Asp Arg Ala Leu Ala Val Val Tyr Ser Val Val Thr Pro Phe	
145 150 155	
CTG	481
Leu	
160	

(2) INFORMATION FOR SEQ ID NO:30:

SUBSTITUTE SHEET

-79-

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 160 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```

Ile Cys Arg Pro Leu His Tyr Pro Thr Leu Met Thr Gln Thr Leu Cys
 1           5           10           15
Ala Lys Ile Ala Thr Gly Cys Trp Leu Gly Gly Leu Ala Gly Pro Val
          20           25           30
Val Glu Ile Ser Leu Val Ser Arg Leu Leu Phe Cys Gly Pro Asn His
          35           40           45
Ile Gln His Ile Phe Cys Asp Phe Pro Pro Val Leu Ser Leu Ala Cys
          50           55           60
Thr Asp Thr Ser Val Asn Val Leu Val Asp Phe Ile Ile Asn Leu Cys
          65           70           75           80
Lys Ile Leu Ala Thr Phe Leu Leu Ile Leu Ser Ser Tyr Leu Gln Ile
          85           90           95
Ile Arg Thr Val Leu Lys Ile Pro Ser Ala Ala Gly Lys Lys Lys Ala
          100          105          110
Phe Ser Thr Cys Ala Ser His Leu Thr Val Val Leu Ile Phe Tyr Gly
          115          120          125
Ser Ile Leu Phe Met Tyr Val Arg Leu Lys Lys Thr Tyr Ser Leu Asp
          130          135          140
Tyr Asp Arg Ala Leu Ala Val Val Tyr Ser Val Val Thr Pro Phe Leu
          145          150          155          160

```

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 481 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: J17

SUBSTITUTE SHEET

-80-

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 2..481

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

A ATC TGC AAC CCA CTG CTT TAT TCC ACC AAA ATG TCC ACA CAA GTC	46
Ile Cys Asn Pro Leu Leu Tyr Ser Thr Lys Met Ser Thr Gln Val	
1 5 10 15	
TGT ATC CAG TTG GTT GCA GGA TCT TAT ATA GGG GGT TTT CTT AAT ACT	94
Cys Ile Gln Leu Val Ala Gly Ser Tyr Ile Gly Gly Phe Leu Asn Thr	
20 25 30	
TGC CTC ATC ATG TTT TAC TTT TTC TCT TTT CTC TTC TGT GGG CCA AAT	142
Cys Leu Ile Met Phe Tyr Phe Phe Ser Phe Leu Phe Cys Gly Pro Asn	
35 40 45	
ATA GTT GAT CAT TTT TTC TGT GAT TTT GCT CCT TTN NTG GAA CTT TCG	190
Ile Val Asp His Phe Phe Cys Asp Phe Ala Pro Xaa Xaa Glu Leu Ser	
50 55 60	
TGC TCT GAT GTG AGT GTC TCT GTA GTT GTT ATG TCA TTT TCT GCT GGC	238
Cys Ser Asp Val Ser Val Ser Val Val Val Met Ser Phe Ser Ala Gly	
65 70 75	
TCA GTT ACT ATG ATC ACA GTG TTT ATC ATA GCC ATC TCC TAT TCT TAC	286
Ser Val Thr Met Ile Thr Val Phe Ile Ile Ala Ile Ser Tyr Ser Tyr	
80 85 90 95	
ATC CTC ATC ACC ATC CTG AAG ATG TCC TCA ACT GAG GGC CGT CAC AAG	334
Ile Leu Ile Thr Ile Leu Lys Met Ser Ser Thr Glu Gly Arg His Lys	
100 105 110	
GCT TTC TCC ACA TGT ACC TCC CAC CTC ACT GCA GTC ACT CTC TAC TAT	382
Ala Phe Ser Thr Cys Thr Ser His Leu Thr Ala Val Thr Leu Tyr Tyr	
115 120 125	
GGC ACC ATT ACC TTC ATT TAT GTG ATG CCC AAG TCC ACA TAC TCT ACA	430
Gly Thr Ile Thr Phe Ile Tyr Val Met Pro Lys Ser Thr Tyr Ser Thr	
130 135 140	
GAC CAG AAC AAG GTG GTG TCT GTG TTT TAC ATG GTG GTG ATC CCA ATG	478
Asp Gln Asn Lys Val Val Ser Val Phe Tyr Met Val Val Ile Pro Met	
145 150 155	
TTG	481
Leu	
160	

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 160 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

SUBSTITUTE SHEET

-81-

```

Ile Cys Asn Pro Leu Leu Tyr Ser Thr Lys Met Ser Thr Gln Val Cys
 1           5           10           15
Ile Gln Leu Val Ala Gly Ser Tyr Ile Gly Gly Phe Leu Asn Thr Cys
           20           25           30
Leu Ile Met Phe Tyr Phe Phe Ser Phe Leu Phe Cys Gly Pro Asn Ile
      35           40           45
Val Asp His Phe Phe Cys Asp Phe Ala Pro Xaa Xaa Glu Leu Ser Cys
      50           55           60
Ser Asp Val Ser Val Ser Val Val Val Met Ser Phe Ser Ala Gly Ser
      65           70           75           80
Val Thr Met Ile Thr Val Phe Ile Ile Ala Ile Ser Tyr Ser Tyr Ile
           85           90           95
Leu Ile Thr Ile Leu Lys Met Ser Ser Thr Glu Gly Arg His Lys Ala
           100           105           110
Phe Ser Thr Cys Thr Ser His Leu Thr Ala Val Thr Leu Tyr Tyr Gly
           115           120           125
Thr Ile Thr Phe Ile Tyr Val Met Pro Lys Ser Thr Tyr Ser Thr Asp
      130           135           140
Gln Asn Lys Val Val Ser Val Phe Tyr Met Val Val Ile Pro Met Leu
      145           150           155           160

```

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 479 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: J19

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..479

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```

T ATC TGC CAC CCT CTG AAG TAC ACA GTT ATC ATG AAT CAC TAT TTT
Ile Cys His Pro Leu Lys Tyr Thr Val Ile Met Asn His Tyr Phe

```

46

SUBSTITUTE SHEET

-82-

1	5	10	15	
TGT GTG ATG CTG CTG CTC TTC TCT GTG TTC GTT AGC ATT GCA CAT GCG				94
Cys Val Met Leu Leu Leu Phe Ser Val Phe Val Ser Ile Ala His Ala	20	25	30	
TTG TTC CAC ATT TTA ATG GTG TTG ATA CTG ACT TTC AGC ACA AAA ACT				142
Leu Phe His Ile Leu Met Val Leu Ile Leu Thr Phe Ser Thr Lys Thr	35	40	45	
GAA ATC CCT CAC TTT TTC TGT GAG CTG GCT CAT ATC ATC AAA CTT ACC				190
Glu Ile Pro His Phe Phe Cys Glu Leu Ala His Ile Ile Lys Leu Thr	50	55	60	
TGT TCC GAT AAT TTT ATC AAC TAT CTG CTG ATA TAC ACA GAG TCT GTC				238
Cys Ser Asp Asn Phe Ile Asn Tyr Leu Leu Ile Tyr Thr Glu Ser Val	65	70	75	
TTA TTT TTT GGT GTT CAT ATT GTA GGG ATC ATT TTG TCT TAT ATT TAC				286
Leu Phe Phe Gly Val His Ile Val Gly Ile Ile Leu Ser Tyr Ile Tyr	80	85	90	95
ACT GTA TCC TCA GTT TTA AGA ATG TCA TTA TTG GGA GGA ATG TAT AAA				334
Thr Val Ser Ser Val Leu Arg Met Ser Leu Leu Gly Gly Met Tyr Lys	100	105	110	
GCC TTT TCA ACA TGT GGA TCT CAT TTG TCG GTT GTC TCT GTT TTA TGG				382
Ala Phe Ser Thr Cys Gly Ser His Leu Ser Val Val Ser Val Leu Trp	115	120	125	
CAC AGG TTT TGG GGT ACA CAT AAG CTC TCC ACT TAC TGA CTC TCC AAG				430
His Arg Phe Trp Gly Thr His Lys Leu Ser Thr Tyr * Leu Ser Lys	130	135	140	
GAA GAC TGT AGT GGC TTC AGT GAT GTA CAC TGT GGT TAC TCA GAT GCT G				479
Glu Asp Cys Ser Gly Phe Ser Asp Val His Cys Gly Tyr Ser Asp Ala	145	150	155	

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 159 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Ile Cys His Pro Leu Lys Tyr Thr Val Ile Met Asn His Tyr Phe Cys															
1				5				10					15		
Val Met Leu Leu Leu Phe Ser Val Phe Val Ser Ile Ala His Ala Leu															
				20				25					30		
Phe His Ile Leu Met Val Leu Ile Leu Thr Phe Ser Thr Lys Thr Glu															
				35				40					45		
Ile Pro His Phe Phe Cys Glu Leu Ala His Ile Ile Lys Leu Thr Cys															
				50				55					60		

SUBSTITUTE SHEET

-83-

Ser Asp Asn Phe Ile Asn Tyr Leu Leu Ile Tyr Thr lu Ser val Leu
 65 70 75 80
 Phe Phe Gly Val His Ile Val Gly Ile Ile Leu Ser Tyr Ile Tyr Thr
 85 90 95
 Val Ser Ser Val Leu Arg Met Ser Leu Leu Gly Gly Met Tyr Lys Ala
 100 105 110
 Phe Ser Thr Cys Gly Ser His Leu Ser Val Val Ser Val Leu Trp His
 115 120 125
 Arg Phe Trp Gly Thr His Lys Leu Ser Thr Tyr • Leu Ser Lys Glu
 130 135 140
 Asp Cys Ser Gly Phe Ser Asp Val His Cys Gly Tyr Ser Asp Ala
 145 150 155

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 481 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: J20

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..481

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

A ATC TGC TAC CCA CTG AGG TAC CTT CTC ATC ATG AGC TGG GTG GTG	46
Ile Cys Tyr Pro Leu Arg Tyr Leu Leu Ile Met Ser Trp Val Val	
1 5 10 15	
TGC ACA GCA CTG TCC GTG GCA ATC TGG GTC ATA GGC TTT TGT GCC TCC	94
Cys Thr Ala Leu Ser Val Ala Ile Trp Val Ile Gly Phe Cys Ala Ser	
20 25 30	
GTT ATA CCT CTC TGC TTC ACG ATC CTC CCA CTC TGT GGT CCT TAC GTC	142
Val Ile Pro Leu Cys Phe Thr Ile Leu Pro Leu Cys Gly Pro Tyr Val	
35 40 45	
GTT GAT TAT CTT TTC TGC GAG CTG CCC ATC CTT CTG CAC CTG TTC TGC	190
Val Asp Tyr Leu Phe Cys Glu Leu Pro Ile Leu Leu His Leu Phe Cys	
50 55 60	
ACA GAT ACA TCT CTC CTG GAG NNN NNN NNN NNN NNN NNN NNN NNN	238
Thr Asp Thr Ser Leu Leu Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
65 70 75	

SUBSTITUTE SHEET

-84-

NNN NNN NNN NNN NNN CCC TTC CTC CTG ATT GTT CTC TCC TAC CTT CGC	286
Xaa Xaa Xaa Xaa Xaa Pro Phe Leu Leu Ile Val Leu Ser Tyr Leu Arg	
80 85 90 95	
ATC CTG GTG GCT GTG ATA AGA ATA GAC TCA GCT GAG GGC AGA AAA AAG	334
Ile Leu Val Ala Val Ile Arg Ile Asp Ser Ala Glu Gly Arg Lys Lys	
100 105 110	
GCC TTT TCA ACT TGT GCT TCA CAC TTG GCT GTG GTG ACC ATC TAC TAT	382
Ala Phe Ser Thr Cys Ala Ser His Leu Ala Val Val Thr Ile Tyr Tyr	
115 120 125	
GGA ACA GGG CTG ATC AGG TAC TTG AGG CCC AAG TCC CTT TAT TCC GCT	430
Gly Thr Gly Leu Ile Arg Tyr Leu Arg Pro Lys Ser Leu Tyr Ser Ala	
130 135 140	
GAG GGA GAC AGA CTG ATC TCT GTG TTC TAT GCA GTC ATT GGC CCT GCA	478
Glu Gly Asp Arg Leu Ile Ser Val Phe Tyr Ala Val Ile Gly Pro Ala	
145 150 155	
CTG	481
Leu	
160	

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 160 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Ile Cys Tyr Pro Leu Arg Tyr Leu Leu Ile Met Ser Trp Val Val Cys	
1 5 10 15	
Thr Ala Leu Ser Val Ala Ile Trp Val Ile Gly Phe Cys Ala Ser Val	
20 25 30	
Ile Pro Leu Cys Phe Thr Ile Leu Pro Leu Cys Gly Pro Tyr Val Val	
35 40 45	
Asp Tyr Leu Phe Cys Glu Leu Pro Ile Leu Leu His Leu Phe Cys Thr	
50 55 60	
Asp Thr Ser Leu Leu Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
65 70 75 80	
Xaa Xaa Xaa Xaa Pro Phe Leu Leu Ile Val Leu Ser Tyr Leu Arg Ile	
85 90 95	
Leu Val Ala Val Ile Arg Ile Asp Ser Ala Glu Gly Arg Lys Lys Ala	
100 105 110	
Phe Ser Thr Cys Ala Ser His Leu Ala Val Val Thr Ile Tyr Tyr Gly	
115 120 125	
Thr Gly Leu Ile Arg Tyr Leu Arg Pro Lys Ser Leu Tyr Ser Ala Glu	
130 135 140	
Gly Asp Arg Leu Ile Ser Val Phe Tyr Ala Val Ile Gly Pro Ala Leu	
145 150 155 160	

SUBSTITUTE SHEET

What is claimed is:

1. An isolated nucleic acid molecule encoding an odorant receptor.
- 5 2. An isolated DNA of claim 1.
3. An isolated cDNA of claim 2.
- 10 4. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 9.
5. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 10.
- 15 6. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 11.
7. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 12.
- 20 8. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 13.
- 25 9. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 14.
10. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 15.
- 30 11. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 16.
12. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequenc shown in Figure 17.
- 35

SUBSTITUTE SHEET

13. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 18.
14. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 19.
15. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 20.
16. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 21.
17. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 22.
18. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 23.
19. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 24.
20. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 25.
21. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 26.
22. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 27.
23. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 28.
24. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 29.

25. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 30.
- 5 26. An isolated cDNA of claim 3 wherein the sequence of the cDNA has the nucleotide sequence shown in Figure 31.
27. An isolated cDNA of claim 3 encoding an insect odorant receptor.
- 10 28. An isolated cDNA of claim 3 encoding a vertebrate odorant receptor.
29. An isolated cDNA of claim 3 encoding a fish odorant receptor.
- 15 30. An isolated cDNA of claim 3 encoding a mammalian odorant receptor.
31. An isolated cDNA of claim 30 wherein the mammalian odorant receptor is a human odorant receptor.
- 20 32. An isolated cDNA of claim 30 wherein the mammalian odorant receptor is a rat, dog or mouse odorant receptor.
- 25 33. An expression vector comprising the cDNA of claim 3 and the sequence elements necessary for replication and expression in a suitable host.
- 30 34. An expression vector comprising the cDNA of any of claims 4-19 and the sequence elements necessary for replication and expression in a suitable host.
35. A purified protein encoding an odorant receptor.

36. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 9.
- 5 37. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 10.
38. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 11.
- 10 39. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 12.
40. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 13.
- 15 41. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 14.
42. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 15.
- 20 43. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 16.
44. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 17.
- 25 45. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 18.
- 30 46. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 19.
- 35 47. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 20.

48. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 21.
- 5 49. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 22.
50. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 23.
- 10 51. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 24.
52. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 25.
- 15 53. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 26.
54. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 27.
- 20 55. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 28.
- 25 56. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 29.
57. A purified protein of claim 35 wherein the amino acid sequence is the sequence in Figure 31.
- 30 58. A purified protein of claim 35 encoding an insect odorant receptor.
59. A purified protein of claim 35 encoding a vertebrate odorant receptor.
- 35

-90-

60. A purified protein of claim 35 encoding a fish odorant receptor.
- 5 61. A purified protein of claim 35 encoding a mammalian odorant receptor.
62. A purified protein of claim 61 wherein the mammalian odorant receptor is a human odorant receptor.
- 10 63. A purified protein of claim 61 wherein the mammalian odorant receptor is a rat, dog or mouse odorant receptor.
- 15 64. A purified protein of claim 35 which has 7 transmembrane regions and whose third cytoplasmic loop from the N-terminus is approximately 17 amino acid long.
- 20 65. A method of transforming cells which comprises transfecting a host cell with a suitable expression vector of claim 33.
- 25 66. A method of transforming cells which comprises transfecting a host cell with a suitable expression vector of claim 34.
67. Cells transformed by the method of claim 65.
- 30 68. Transformed cells of claim 67 wherein the cells are olfactory cells.
69. Transformed cells of claim 67 wherein the cells are non-olfactory cells.
- 35

- 5 70. A method of identifying a desired odorant ligand comprising contacting transformed non-olfactory cells of claim 69, expressing a known odorant receptor with a series of odorant ligands and determining which ligands bind to the receptors present on the non-olfactory cells.
- 10 71. A method of identifying a desired odorant receptor comprising contacting a series of transformed non-olfactory cells of claim 69 with a known odorant ligand and determining which odorant receptor binds with the odorant ligand.
- 15 72. A method of detecting an odor which comprises:
- a) identifying a odorant receptor which binds the desired odorant ligand by the method of claim 71 and;
- 20 b) imbedding the receptor in a membrane such that when the odorant ligand binds with the receptor identified in a) above, a detectable signal is produced.
- 25 73. A method of claim 72 wherein the desired odorant is a pheromone.
- 30 74. A method of claim 72 wherein the desired odorant ligand is the vapors emanating from cocaine, marijuana, heroin, hashish, or angel dust.
- 35 75. A method of claim 72 wherein the desired odorant ligand is the vapors emanating from gasoline, natural gas or alcohol.

-92-

76. A method of claim 72 wherein the desired odorant ligand is the vapors emanating from decayed human flesh.
- 5 77. A method of claim 72 wherein the desired odorant ligand is the vapors emanating from explosives, plastic explosives, firearms, or gun powder.
78. A method of claim 72 wherein the desired odorant ligand is toxic fumes, noxious fumes or dangerous fumes.
- 10 79. A method of claim 72 wherein the membrane is a cell membrane.
80. A method of claim 72 wherein the membrane is an olfactory cell membrane.
- 15 81. A method of claim 72 wherein the membrane is a synthetic membrane.
- 20 82. A method of claim 72 wherein the detectable signal is a color change, phosphorescence, or radioactivity.
83. A method of quantifying the amount of an odorant ligand present in a sample which comprises the method of claim 72 wherein the detectable signal is quantified.
- 25 84. A method of developing fragrances which comprises identifying a desired odorant receptor by the method of claim 71 then contacting non-olfactory cells, which have been transfected with an expression vector containing the cDNA of the desired odorant receptor such that the receptor is expressed upon the surface of the non-olfactory cell, with a series of compounds to determine which ones bind with the receptor.
- 30
- 35

SUBSTITUTE SHEET

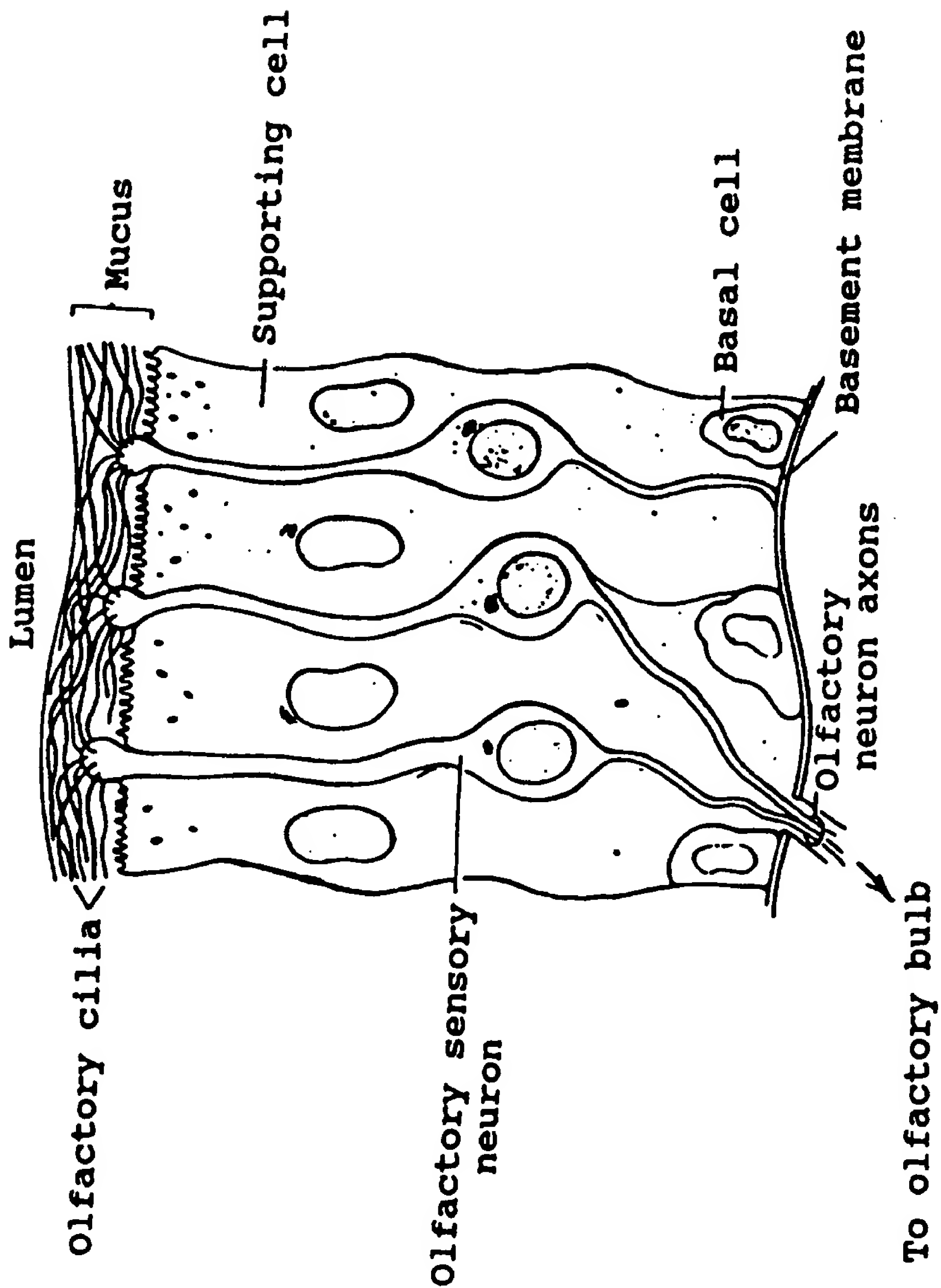
- 5 85. A method of identifying an odorant fingerprint which comprises contacting a series of cells, which have been transformed such that each express a known odorant receptor, with a desired sample and determining the type and quantity of the odorant ligands present in the sample.
- 10 86. A method of identifying odorant ligands which inhibit the activity of a desired odorant receptor which comprises contacting the desired odorant receptor with a series of compounds and determining which compounds inhibit the odorant ligand - odorant receptor interaction.
- 15 87. A method of identifying appetite suppressant compounds which comprises identifying odorant ligands by the method of claim 86 wherein the desired odorant receptor is that which is associated with the perception of food.
- 20 88. A method of controlling appetite in a subject which comprises contacting the olfactory epithelium of the subject with the odorant ligands identified by the method of claim 87.
- 25 89. A nasal spray, to control appetite comprising the compounds identified by the method of claim 87 in a suitable carrier.
- 30 90. A method of trapping odors which comprises contacting a membrane which contains multiples of the desired odorant receptor, with a sample such that the desired odorant ligand is absorbed by the binding of the odorant ligand to the odorant receptor.
- 35

-94-

91. An odor trap employing the method of claim 90.
92. A method of controlling pest populations which comprises identifying odorant ligands by the method of claim 70 which are alarm odorant ligands and spraying the desired area with the identified odorant ligands.
93. A method of controlling a pest population which comprises identifying odorant ligands by the method of claim 70 which interfere with the interaction between the odorant ligands and the odorant receptors which are associated with fertility.
94. A method of claim 92 or 93 wherein the pest population is a population of insects.
95. A method of claim 92 or 93 wherein the pest population is a population of rodents.
96. A method of claim 95 wherein the population of rodents is a population of mice or rats.
97. A method of promoting fertility which comprises employing the method of claim 70 to identify odorant ligands which interact with the odorant receptors associated with fertility and administering the identified odorant ligands to a subject.
98. A method of inhibiting fertility which comprises employing the method of claim 70 to identifying odorant ligands which inhibit the interaction between the odorant ligands and the odorant receptors associated with fertility and administering the identified odorant ligands to a subject.

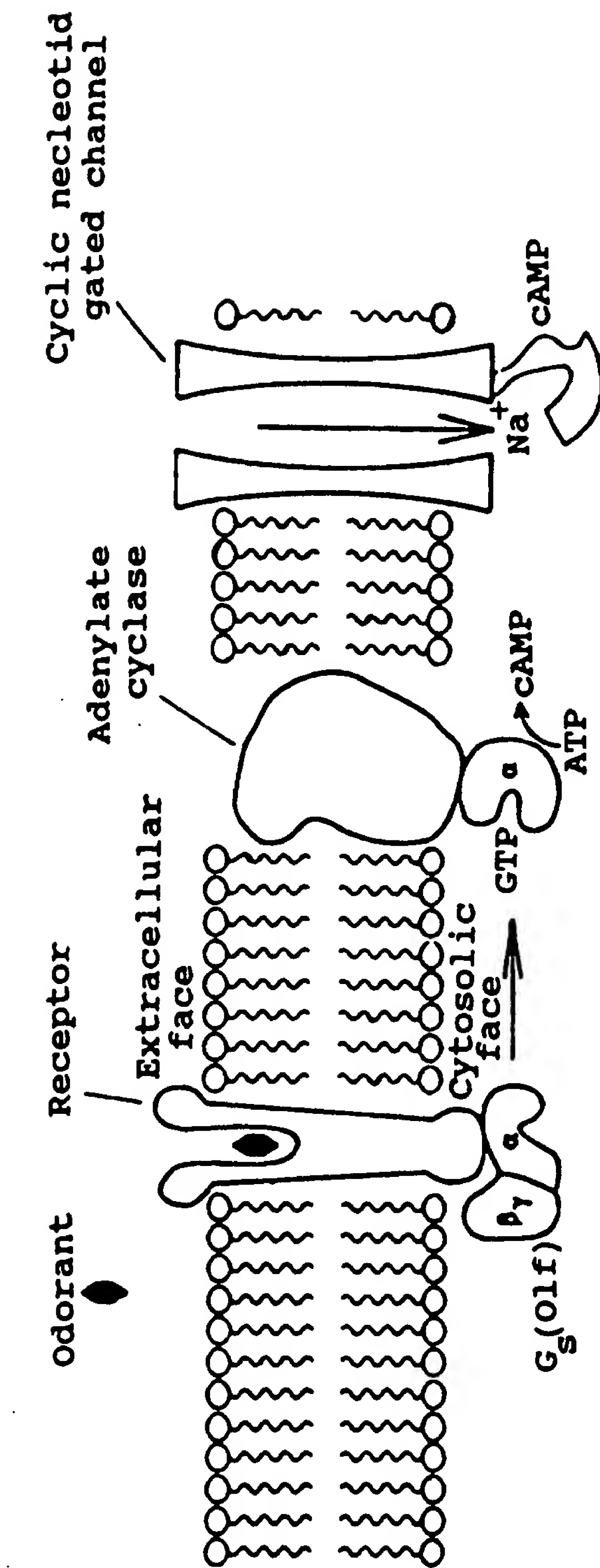
SUBSTITUTE SHEET

Figure 1A

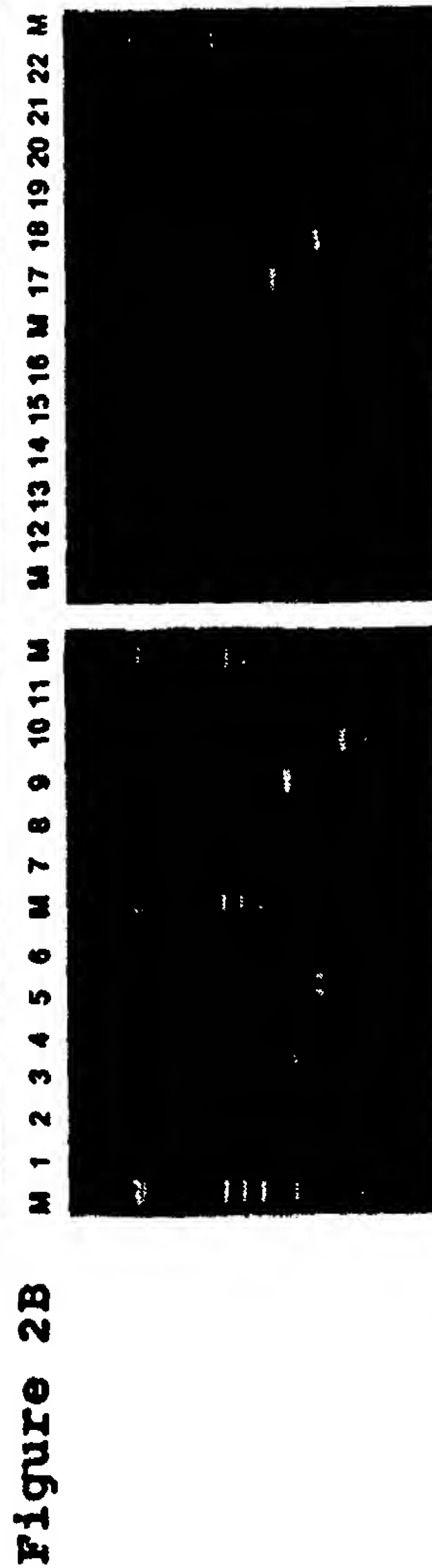
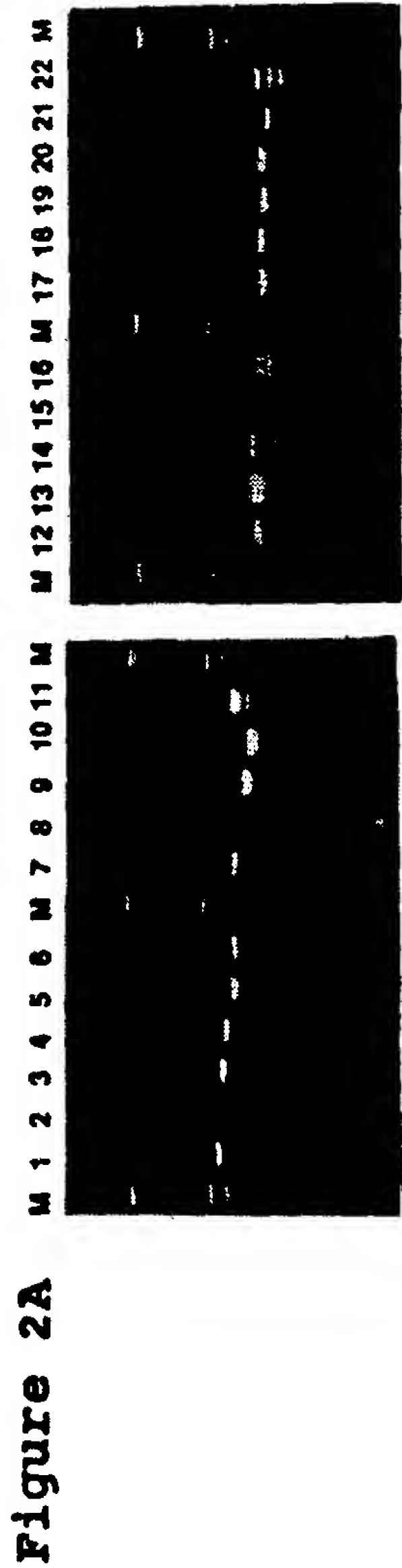


SUBSTITUTE SHEET

Figure 1B



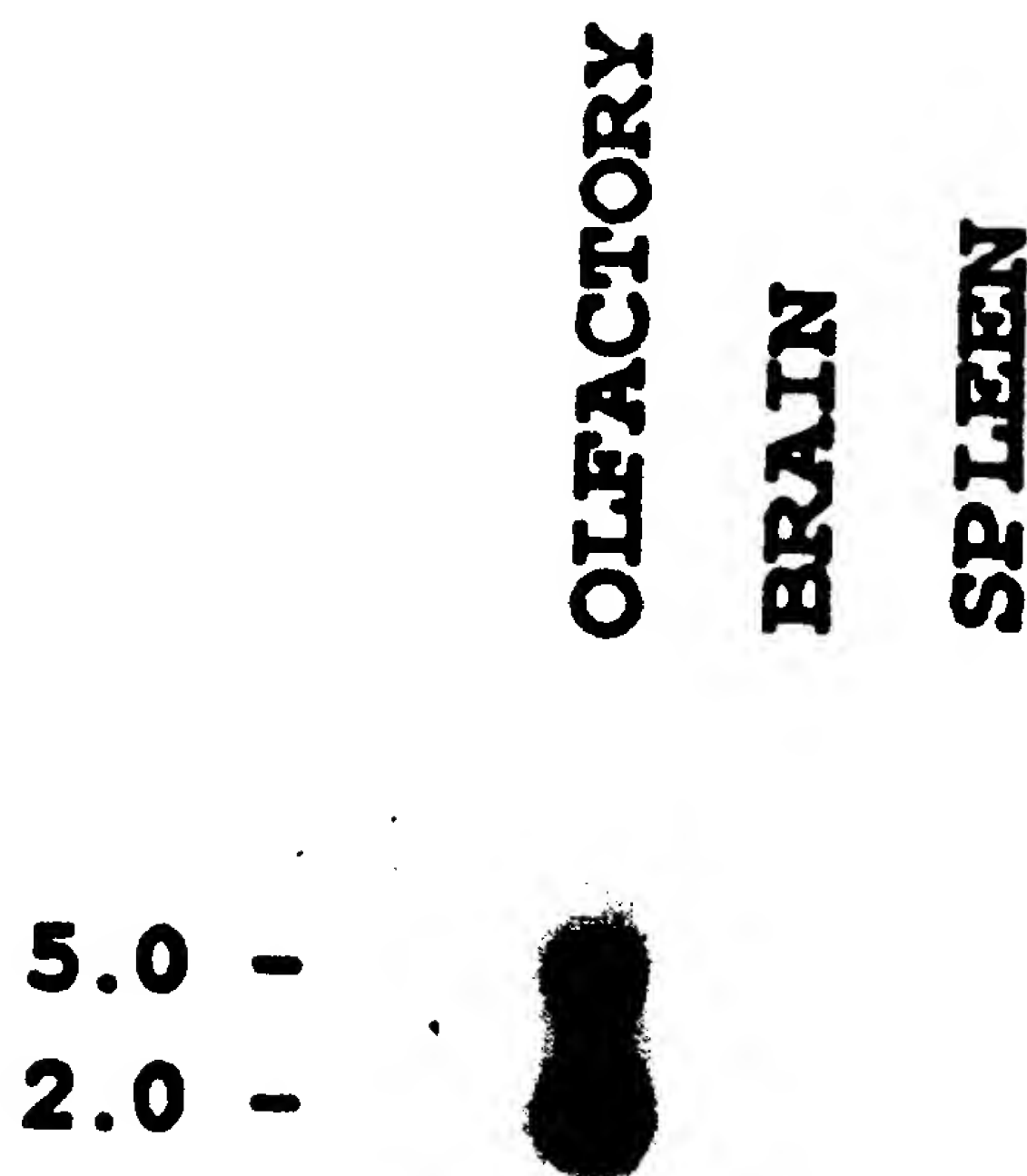
SUBSTITUTE SHEET



SUBSTITUTE SHEET

4 / 98

Figur 3



SUBSTITUTE SHEET

Figur 4A

F3		N	D	S	S	N	R	T	R	V	S	E	11
F5		N	S	S	T	N	Q	S	S	V	T	E	11
F6	N A W	S	T	G	Q	N	L	S	T	P	G	P	14
F12		N	E	S	G	N	S	T	R	R	F	S	12
I3		N	N	-	-	N	Q	T	F	I	T	Q	9
I7		N	E	R	R	N	H	S	G	R	V	S	12
I8		N	N	-	-	N	K	T	V	I	T	H	9
I8		N	T	R	R	N	Q	T	A	I	S	Q	11
I14		N	T	G	N	N	Q	T	L	I	L	E	11
I15		N	T	E	E	N	Q	T	V	I	S	Q	11

F3	F	L	L	L	G	F	V	E	N	K	D	L	Q	P	25
F5	F	L	L	L	G	L	S	R	Q	P	Q	Q	Q	Q	25
F6	F	I	L	L	G	F	P	G	P	R	S	M	R	I	28
F12	F	F	L	L	G	F	T	E	N	P	Q	L	H	F	26
I3	F	L	L	L	G	L	P	I	P	E	E	H	Q	H	23
I7	F	V	L	L	G	F	P	A	P	A	P	L	R	V	26
I8	F	L	L	L	G	L	P	I	P	P	E	H	Q	Q	23
I9	F	F	L	L	G	L	P	F	P	P	E	Y	Q	H	25
I14	F	L	L	L	G	L	P	I	P	S	E	Y	H	L	25
I15	F	L	L	L	F	L	P	I	P	S	E	H	Q	H	25

6 / 98

Figur 4B

	<u>I</u>	
F3	L I Y G L F L S N Y L V T V	39
F5	L L F L L F L I N Y L A T V	39
F6	G L F L L F L V N Y L L T V	42
F12	L I F A L F L S N Y L V T V	40
I3	L F Y A L F L V N Y L T T I	37
I7	L L F F L S L L X Y V L V L	40
I8	L F F A L F L I N Y L T T F	37
I9	L F Y A L F L A N Y L T T L	39
I14	L F Y A L F L A N Y L T I I	29
I15	V F Y A L F L S N Y L T T V	39

	<u>I</u>	
F3	I G N I S I I V A I I S D P	53
F5	L G N L L I I L A I G T D S	53
F6	V G N L A I I S L V G A H R	56
F12	L G N L L I I M A I I T Q S	54
I3	L G N L L I I V L V Q L D S	51
I7	T E N M L I I I A I R N H P	54
I8	L G N L L I V V L V Q L D S	51
I9	L G N L I I I I L I L L D S	53
I14	L G N L L I I V L V R L D S	53
I15	L G N L I I I I L I H L D S	53

Figur 4C

	<u>II</u>												
F3	C	L	H	T	P	N	Y	F	F	L	S	N	L S 67
F5	R	L	H	T	P	N	Y	F	F	L	S	N	L S 67
F6	C	L	Q	T	P	N	Y	F	F	L	C	N	L S 70
F12	H	L	H	T	P	N	Y	F	F	L	A	N	L S 68
I3	Q	L	H	T	P	N	Y	L	F	L	S	N	L S 65
I7	T	L	H	K	P	N	Y	F	F	L	A	N	M S 68
I8	H	L	H	T	P	N	Y	L	F	L	S	N	L S 65
I9	H	L	H	T	P	N	Y	L	F	L	S	N	L S 67
I14	H	L	H	M	P	N	Y	L	F	L	S	N	L S 67
I15	H	L	H	T	P	N	Y	L	F	L	S	N	L S 67

	<u>II</u>												
F3	F	V	D	I	C	F	I	S	T	T	V	P	K M 81
F5	F	V	D	V	C	F	S	S	T	T	V	P	K V 81
F6	F	L	E	I	W	F	T	T	A	C	V	P	K T 84
F12	F	V	D	I	C	F	T	S	T	T	I	P	K M 82
I3	F	S	D	L	C	F	S	S	V	T	M	P	K L 79
I7	F	L	E	I	W	Y	V	T	V	T	I	P	K M 82
I8	F	S	D	L	C	F	S	S	V	T	M	L	K L 79
I9	F	A	D	L	C	F	S	S	V	T	M	P	K L 67
I14	F	S	D	L	C	F	S	S	V	T	M	P	K L 67
I15	F	S	D	L	C	F	S	S	V	T	M	P	K L 67

8 / 98

Figur 4D

F3	L	-	-	-	-	V	N	I	Q	T	Q	N	N	V	91
F5	L	-	-	-	-	A	N	H	I	L	G	S	Q	A	91
F6	L	-	-	-	-	A	T	F	A	P	R	G	G	V	94
F12	L	-	-	-	-	V	N	I	Y	T	Q	S	K	S	92
I3	L	-	-	-	-	Q	N	M	R	S	Q	K	T	S	89
I7	L	A	G	F	I	G	S	K	E	N	H	G	Q	L	96
I8	L	-	-	-	-	Q	N	I	Q	S	Q	V	P	S	89
I9	L	-	-	-	-	Q	N	M	Q	S	Q	V	P	S	91
I14	L	-	-	-	-	Q	N	M	Q	S	Q	V	P	S	91
I15	L	-	-	-	-	Q	N	M	Q	S	Q	V	P	S	91

									<u>III</u>						
F3	I	T	Y	A	G	C	I	T	Q	I	Y	F	F	L	105
F5	I	S	F	S	G	C	L	T	Q	L	Y	F	L	A	105
F6	I	S	L	A	G	C	A	T	Q	M	Y	F	V	F	108
F12	I	T	Y	E	D	C	I	S	Q	M	C	V	F	L	106
I3	I	P	Y	G	G	C	L	A	Q	T	Y	F	F	M	103
I7	I	S	F	E	A	C	M	T	Q	L	Y	F	F	L	110
I8	I	S	Y	A	G	C	L	T	Q	I	F	F	F	L	103
I9	I	P	Y	A	G	C	L	A	Q	I	Y	F	F	L	105
I14	I	S	Y	T	G	C	L	T	Q	L	Y	F	F	M	105
I15	I	P	F	A	G	C	L	T	Q	L	Y	F	Y	L	105

Figur 4E

	<u>III</u>	
F3	L F V E L D N F L L T I N A	119
F5	V F G N M D N F L L A V N S	119
F6	S L G C T E Y F L L A V N A	122
F12	V F A I L G N F L L A V N A	120
I3	V F G D M E S F L L V A N A	117
I7	G L G C T E C V L L A V N A	124
I8	L F G Y L G N F L L V A N A	117
I9	F F G D L G N F L L V A N A	119
I14	V F G D M E S F L L V V N A	119
I15	Y F A D L E S F L L V A N A	119

	<u>III</u>	
F3	Y D R Y V A I C H P M H Y T	133
F5	Y D R F V A I C H P L H Y T	133
F6	Y D R Y L A I C L P L R Y G	136
F12	Y D R Y V A X C H P L C Y T	134
I3	Y D R Y V A I C F P L H Y T	131
I7	Y D R Y V A I C H P L H Y P	138
I8	Y D R Y V A I C F P L H Y T	131
I9	Y D R Y V A I C F P L H Y M	133
I14	Y D R Y V A I C F P L R Y T	133
I15	Y D R Y V A I C F P L H Y M	133

Figure 4F

	<u>IV</u>														
F3	V	I	N	N	Y	K	L	C	G	F	L	V	L	V	147
F5	T	K	N	T	R	Q	L	C	V	L	L	V	V	G	147
F6	G	I	N	T	P	G	L	A	M	R	L	A	L	G	150
F12	V	I	V	N	H	R	L	C	I	L	L	L	L	L	148
I3	S	I	N	S	P	K	L	C	T	C	L	V	L	L	145
I7	V	I	V	S	S	R	L	C	V	Q	M	A	A	G	152
I8	N	I	N	S	H	K	L	C	T	C	L	L	L	V	145
I9	S	I	N	S	P	K	L	C	V	S	L	V	V	L	147
I14	T	I	N	S	T	K	F	C	A	S	L	V	L	L	147
I15	S	I	N	S	P	K	L	C	V	S	L	V	V	L	147

	<u>IV</u>														
F3	S	N	I	V	S	V	L	H	A	L	F	Q	S	L	161
F5	S	N	V	V	A	N	M	N	C	L	L	H	I	L	161
F6	S	N	L	C	G	F	S	A	I	T	V	P	A	T	164
F12	S	N	V	I	S	I	F	H	A	F	I	Q	S	L	162
I3	L	N	M	L	T	T	S	H	A	M	M	H	T	L	159
I7	S	N	A	G	G	F	G	I	S	M	V	K	V	F	166
I8	F	N	I	M	T	S	S	H	A	M	M	H	T	L	159
I9	S	N	V	L	T	T	F	H	A	M	L	H	T	L	161
I14	L	N	M	L	T	M	T	H	A	L	L	H	T	L	161
I15	S	N	V	L	T	T	F	H	A	M	L	H	T	L	161

11/98

Figur 4G

F3	M	M	L	A	L	P	F	C	T	H	L	E	I	P	175
F5	L	M	A	R	K	S	F	C	A	D	N	M	I	P	175
F6	L	I	A	R	L	S	F	C	G	S	R	V	I	N	178
F12	I	V	L	Q	L	T	F	C	G	D	V	K	I	P	176
I3	L	A	A	R	L	S	F	C	E	N	N	V	V	L	173
I7	L	I	S	R	L	S	Y	C	G	P	N	T	I	N	180
I8	L	A	A	R	L	S	F	C	E	N	N	V	L	L	173
I9	L	M	A	R	L	S	F	C	E	D	S	V	I	P	175
I14	L	I	A	R	L	S	F	C	E	K	N	V	I	L	175
I15	L	M	A	R	L	S	F	C	A	D	N	M	I	P	175

F3	H	Y	F	C	E	P	N	Q	V	I	Q	L	T	C	189
F5	H	F	F	C	D	G	T	P	L	L	K	L	S	C	189
F6	H	F	F	C	D	I	S	P	W	I	V	L	S	C	192
F12	H	F	F	C	E	L	N	Q	L	S	Q	L	T	C	190
I3	N	F	F	C	D	L	F	V	L	L	K	L	A	C	187
I7	H	F	F	C	D	V	S	P	L	L	N	L	S	C	194
I8	N	F	F	C	D	L	F	V	L	L	K	L	A	C	187
I9	H	Y	F	C	D	M	S	T	L	L	K	V	A	C	189
I14	H	F	F	C	D	I	S	A	L	L	K	L	S	C	189
I15	H	F	F	C	D	I	S	P	L	L	K	L	S	C	189

SUBSTITUTE SHEET

Figur 4H

								<u>V</u>							
F3	S	D	A	F	L	N	D	L	V	I	Y	F	T	L	203
F5	S	D	T	H	L	N	E	L	M	I	L	T	E	G	203
F6	T	D	T	Q	V	V	E	L	V	S	F	G	I	A	206
F12	S	D	N	F	P	S	H	L	I	M	N	L	V	P	204
I3	S	D	T	Y	I	N	E	L	M	I	F	I	M	S	201
I7	T	D	M	S	T	A	E	L	T	D	F	V	L	A	208
I8	S	D	T	Y	V	N	E	L	M	I	H	I	M	G	201
I9	S	D	T	H	D	N	E	L	A	I	F	I	L	G	203
I14	S	D	I	Y	V	N	E	L	M	I	Y	I	L	G	203
I15	S	D	T	H	V	N	E	L	V	I	F	V	M	G	203

							<u>V</u>							
F3	V	L	L	A	T	V	P	L	A	G	I	F	Y S	217
F5	A	V	V	M	V	T	P	F	V	C	I	L	I S	217
F6	F	C	V	I	L	G	S	C	G	I	T	L	V S	220
F12	V	M	L	A	A	I	S	F	S	G	I	L	Y S	218
I3	T	L	L	I	I	I	P	F	F	L	I	V	M S	215
I7	I	F	I	L	L	G	P	L	S	V	T	G	A S	222
I8	V	I	I	I	V	I	P	F	V	L	I	V	I S	215
I9	G	P	I	V	V	L	P	F	L	L	I	I	V S	203
I14	G	L	I	I	I	I	P	F	L	L	I	V	M S	203
I15	G	L	V	I	V	I	P	F	V	L	I	I	V S	203

13/98

Figur 4I

	<u>V</u>														
F3	Y	F	K	I	V	S	S	I	C	A	I	S	S	V	231
F5	Y	I	H	I	T	C	A	V	L	R	V	S	S	P	231
F6	Y	A	Y	I	I	T	T	I	I	K	I	P	S	A	234
F12	Y	F	K	I	V	S	S	I	H	S	I	S	T	V	232
I3	Y	A	R	I	I	S	S	I	L	K	V	P	S	T	229
I7	Y	M	A	I	T	G	A	V	M	R	I	P	S	A	236
I8	Y	A	K	I	I	S	S	I	L	K	V	P	S	T	229
I9	Y	A	R	I	V	S	S	I	F	K	V	P	S	S	231
I14	Y	V	R	I	F	F	S	I	L	K	F	P	S	I	231
I15	Y	A	R	V	V	A	S	I	L	K	V	P	S	V	231

	<u>VI</u>														
F3	H	G	K	Y	K	A	F	S	T	C	A	S	H	L	245
F5	R	G	G	W	K	S	F	S	T	C	G	S	H	L	245
F6	R	G	R	H	R	A	F	S	T	C	S	S	H	L	248
F12	Q	G	K	Y	K	A	F	S	T	C	A	S	H	L	246
I3	Q	G	I	C	K	V	F	S	T	C	G	S	H	L	243
I7	A	G	R	H	K	A	F	S	T	C	A	S	H	L	250
I8	Q	S	I	H	K	V	F	S	T	C	G	S	H	L	243
I9	Q	S	I	H	K	A	F	S	T	C	G	S	H	L	245
I14	Q	D	I	Y	K	V	F	S	T	C	G	S	H	L	245
I15	R	G	I	H	K	I	F	S	T	C	G	S	H	L	245

14/98

Figur 4J

	<u>VI</u>	
F3	S V V S L F Y C T G L G V Y	259
F5	A V V C L F Y G T V I A V Y	259
F6	T V V L I W Y G S T I F L H	262
F12	S I V S L F Y S T G L G V Y	260
I3	S V V S L F Y G T I I G L Y	257
I7	T V V I I F Y A A S I F I Y	264
I8	S V V S L F Y G T I I G L Y	257
I9	S V V S L F Y G T V I G L Y	259
I14	S V V T L F Y G T I F G I Y	259
I15	S V V S L F Y G T I I G L Y	259

	<u>VI</u>	<u>VII</u>	
F3	L S S A A N N S S Q A S A T		273
F5	F N P S S S H L A G R D M A		273
F6	V R T S V E S S L D L T K A		276
F12	V S S A V V Q S S H S A A S		274
I3	L C P A G N N S T V K E M V		271
I7	A R P K A L S A F D T N K L		278
I8	L C P S G D N F S L K G S A		271
I9	L C P S A N N S T V K E T V		273
I14	L C P S G N N S T V K E I A		273
I15	L C P S A N N S T V K E T V		273

15/98

Figur 4K

	<u>VII</u>														
F3	A	S	V	M	Y	T	V	V	T	P	M	V	N	P	287
F5	A	A	V	M	Y	A	V	V	T	P	M	L	N	P	287
F6	I	T	V	L	N	T	I	V	T	P	V	L	N	P	290
F12	A	S	V	M	Y	T	V	V	T	P	M	L	N	P	288
I3	M	A	M	M	Y	T	V	V	T	P	M	L	N	P	285
I7	V	S	V	L	Y	A	V	I	V	P	L	F	N	P	292
I8	M	A	M	M	Y	T	V	V	T	P	M	L	N	P	285
I9	M	S	L	M	Y	T	M	V	T	P	M	L	N	P	287
I14	M	A	M	M	Y	T	V	V	T	P	M	L	N	P	287
I15	M	A	M	M	Y	T	V	V	T	P	M	L	N	P	287

	<u>VII</u>														
F3	F	I	Y	S	L	R	N	K	D	V	K	S	V	L	301
F5	F	I	Y	S	L	R	N	S	D	M	K	A	A	L	301
F6	F	I	Y	T	L	R	N	K	D	V	K	E	A	L	304
F12	F	I	Y	S	L	R	N	K	D	V	K	R	A	L	302
I3	F	I	Y	S	L	R	N	R	D	M	K	R	A	L	299
I7	I	I	Y	C	L	R	N	Q	D	V	K	R	A	L	306
I8	F	I	Y	S	L	R	N	R	D	M	K	Q	A	L	299
I9	F	I	Y	S	L	R	N	R	D	I	K	D	A	L	301
I14	F	I	Y	S	L	R	N	R	D	M	K	R	A	L	301
I15	F	I	Y	S	L	R	N	R	D	M	K	E	A	L	301

SUBSTITUTE SHEET

Figur 4L

F3	K	K	T	L	C	E	E	V	I	R	S	P	P	S	315
F5	R	K	V	L	A	M	R	F	P	S	K	Q	-		313
F6	R	R	T	V	K	G	K	-							311
F12	E	R	L	L	E	G	N	C	K	V	H	H	W	T	316
I3	I	R	V	I	C	S	M	K	I	T	L	-			310
I7	R	R	T	L	H	L	A	Q	D	Q	E	A	N	T	320
I8	I	R	V	T	C	S	K	K	I	S	L	P	W	-	312
I9	E	K	I	M	C	K	K	Q	I	P	S	F	L	-	314
I14	I	R	V	I	C	T	K	K	I	S	L	-			312
I15	I	R	V	L	C	K	K	K	I	T	F	C	L	-	314

F3	L	L	H	F	F	L	V	L	C	H	L	P	C	F	329
F5															
F6															
F12	G	-													317
I3															
I7	N	K	G	S	K	I	G	-							327
I8															
I9															
I14															
I15															

SUBSTITUTE SHEET

17/98

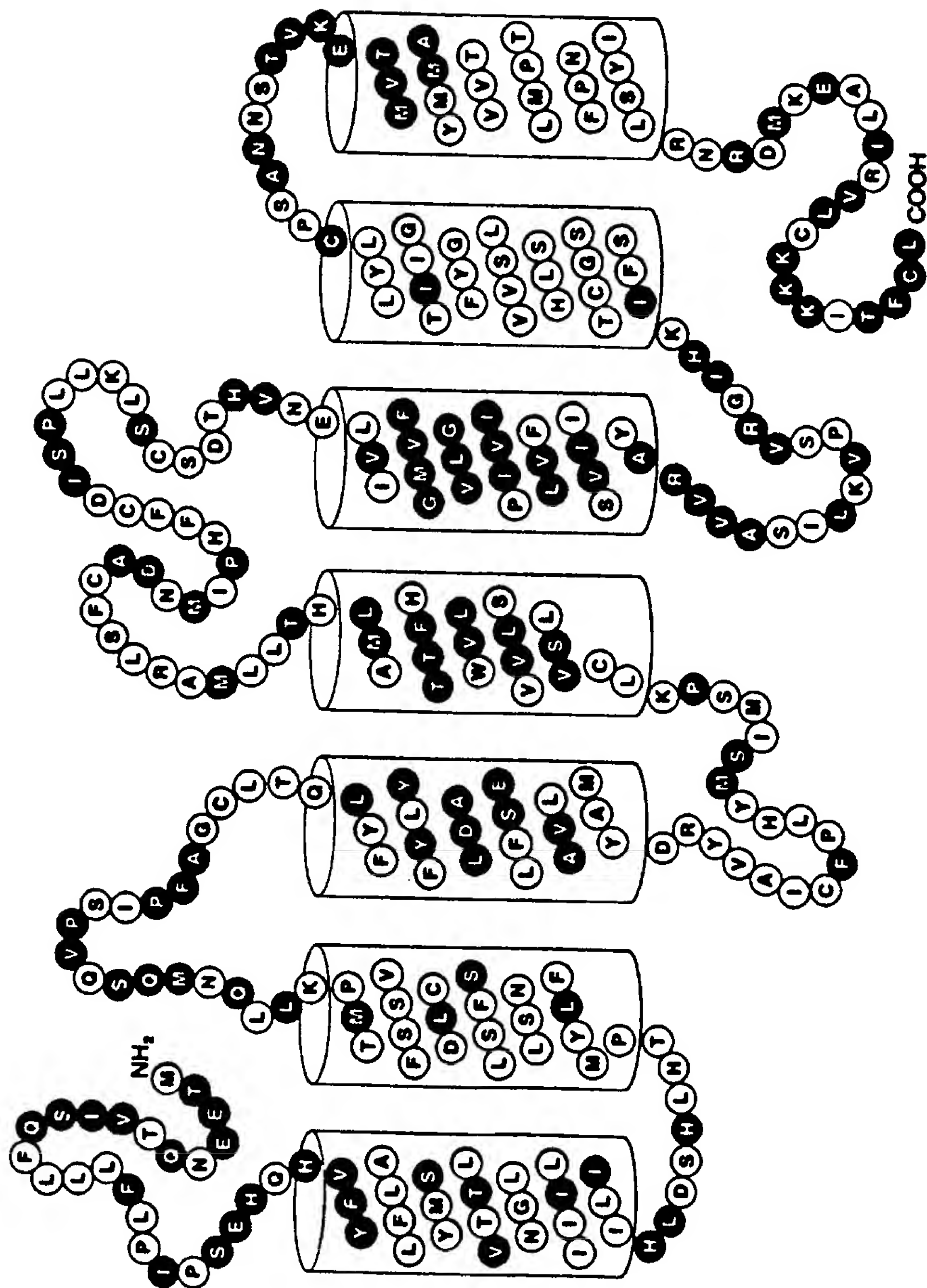
Figure 4M

F3 I F C Y -
F5
F6
F12
I3
I7
I8
I9
I14
I15

333

SUBSTITUTE SHEET

Figure 5



19/98

Figure 6A(1)

					V														
F2	R	V	N	E	V	V	I	F	I	V	V	S	L	F					
F3	F	L	N	D	L	V	I	Y	F	T	L	V	L	L					
F5	H	L	N	E	L	M	I	L	T	E	G	A	V	V					
F6	Q	V	V	E	L	V	S	F	G	I	A	F	C	V					
F7	H	V	N	E	L	V	I	F	V	M	G	G	I	I					
F8	F	P	S	H	L	T	M	H	L	V	P	V	I	L					
F12	F	P	S	H	L	I	M	N	L	V	P	V	M	L					
F13	F	P	S	H	L	I	M	N	L	V	P	V	M	L					
F23	F	L	N	D	V	I	M	Y	F	A	L	V	L	L					
F24	H	E	I	E	M	I	I	L	V	L	A	A	F	N					
I3	Y	I	N	E	L	M	I	F	I	M	S	T	L	L					
I7	S	T	A	E	L	T	D	F	V	L	A	I	F	I					
I8	Y	V	N	E	L	M	I	H	I	M	G	V	I	I					
I9	H	D	N	E	L	A	I	F	I	L	G	G	P	I					
I11	H	L	N	E	L	M	I	L	T	E	G	A	V	V					
I12	F	P	S	H	L	I	M	N	L	V	P	V	M	L					
I14	Y	V	N	E	L	M	I	Y	I	L	G	G	L	I					
I15	H	V	N	E	L	V	I	F	V	M	G	G	L	V					

SUBSTITUTE SHEET

20/98

Figure 6A(2)

	V												
F2	L	V	L	P	F	A	L	I	I	M	S	Y	V R
F3	A	T	V	P	L	A	G	I	F	Y	S	Y	F K
F5	M	V	T	P	F	V	C	I	L	I	S	Y	I H
F6	I	H	G	S	C	G	I	T	L	V	S	Y	A Y
F7	L	V	I	P	F	V	L	I	I	V	S	Y	V R
F8	A	A	I	S	L	S	G	I	L	Y	S	Y	F K
F12	A	A	I	S	F	S	G	I	L	Y	S	Y	F K
F13	A	A	I	S	F	S	G	I	L	Y	S	Y	F K
F23	A	V	V	P	L	L	G	I	L	Y	S	Y	S K
F24	L	I	S	S	L	L	V	V	L	V	S	Y	L F
I3	I	I	I	P	F	F	L	I	V	M	S	Y	A R
I7	L	L	G	P	L	S	V	T	G	A	S	Y	M A
I8	I	V	I	P	F	V	L	I	V	I	S	Y	A K
I9	V	V	L	P	F	L	L	I	I	V	S	Y	A R
I11	M	V	T	P	F	V	C	I	L	I	S	Y	I H
I12	G	A	I	S	L	S	G	I	L	Y	S	Y	F K
I14	I	I	I	P	F	L	L	I	V	M	S	Y	V R
I15	I	V	I	P	F	V	L	I	I	V	S	Y	A R

21 / 98

Figure 6A(3)

F2	I	V	S	S	I	L	K	V	P	S	S	Q	G	I
F3	I	V	S	S	I	C	A	I	S	S	V	H	G	K
F5	I	T	C	A	V	L	R	V	S	S	P	R	G	G
F6	I	I	T	T	I	I	K	I	P	S	A	R	G	R
F7	I	V	S	S	I	L	K	V	P	S	A	R	G	I
F8	I	V	S	S	I	R	S	M	S	S	V	Q	G	K
F12	I	V	S	S	I	H	S	I	S	T	V	Q	G	K
F13	I	V	S	S	I	R	S	V	S	S	V	K	G	K
F23	I	V	S	S	I	R	A	I	S	T	V	Q	G	K
F24	I	L	I	A	I	L	R	M	N	S	A	E	G	R
I3	I	I	S	S	I	L	K	V	P	S	T	Q	G	I
I7	I	T	G	A	V	M	R	I	P	S	A	A	G	R
I8	I	I	S	S	I	L	K	V	P	S	T	Q	S	I
I9	I	V	S	S	I	F	K	V	P	S	S	Q	S	I
I11	I	T	W	A	V	L	R	V	S	S	P	R	G	G
I12	I	V	S	S	V	R	S	I	S	S	V	Q	G	K
I14	I	F	F	S	I	L	K	F	P	S	I	Z	D	I
I15	V	V	A	S	I	L	K	V	P	S	V	R	G	I

22/98

Figure 6A(4)

F2	Y	K
F3	Y	K
F5	W	K
F6	H	R
F7	R	K
F8	Y	K
F12	Y	K
F13	Y	K
F23	Y	K
F24	R	K
I3	C	K
I7	H	K
I8	H	K
I9	H	K
I11	W	K
I12	H	K
I14	Y	K
I15	H	K

SUBSTITUTE SHEET

23/98

Figure 6B

					V									
F12	F	P	S	H	L	I	N	N	L	V	P	V	M	L
F13	F	P	S	H	L	I	N	N	L	V	P	V	M	L
F8	F	P	S	H	L	T	N	H	L	V	P	V	I	L
I12	F	P	S	H	L	I	N	N	L	V	P	V	M	L
F23	F	L	N	D	V	I	N	Y	F	A	L	V	L	L
F3	F	L	N	D	L	V	I	Y	F	T	L	V	L	L

					V									
F12	A	A	I	S	F	S	G	I	L	Y	S	Y	F	K
F13	A	A	I	S	F	S	G	I	L	Y	S	Y	F	K
F8	A	A	I	S	L	S	G	I	L	Y	S	Y	F	K
I12	G	A	I	S	L	S	G	I	L	Y	S	Y	F	K
F23	A	V	V	P	L	L	G	I	L	Y	S	Y	S	K
F3	A	T	V	P	L	A	G	I	F	Y	S	Y	F	K

SUBSTITUTE SHEET.

24/98

Figure 6B (Continued)

F12	I	V	S	S	I	H	S	I	S	T	V	Q	G	K
F13	I	V	S	S	I	R	S	V	S	S	V	K	G	K
F8	I	V	S	S	I	R	S	M	S	S	V	Q	G	K
I12	I	V	S	S	V	R	S	I	S	S	V	Q	G	K
F23	I	V	S	S	I	R	A	I	S	T	V	Q	G	K
F3	I	V	S	S	I	C	A	I	S	S	S	H	G	K

F12	Y	K
F13	Y	K
F8	Y	K
I12	H	K
F23	Y	K
F3	Y	K

SUBSTITUTE SHEET

25 / 98

Figure 6C

					V								
F7	H	V	N	E	L	V	I	F	V	M	G	G	I
I15	H	V	N	E	L	V	I	F	V	M	G	G	L
I3	Y	I	N	E	L	M	I	F	I	N	S	T	L
I8	Y	V	N	E	L	M	I	H	I	N	G	V	I
I9	H	D	N	E	L	A	I	F	I	L	G	G	P
I14	Y	V	N	E	L	M	I	Y	I	L	G	G	L

					V								
F7	L	V	I	P	F	V	L	I	I	V	S	Y	V
I15	I	V	I	P	F	V	L	I	I	V	S	Y	A
I3	I	I	I	P	F	F	L	I	V	M	S	Y	A
I8	I	V	I	P	F	V	L	I	V	I	S	Y	A
I9	V	V	L	P	F	L	L	I	I	V	S	Y	A
I14	I	I	I	P	F	L	L	I	V	M	S	Y	V

SUBSTITUTE SHEET

Figure 6C (Continued)

F7	I	V	S	S	I	L	K	V	P	S	A	R	G	I
I15	V	V	A	S	I	L	K	V	P	S	V	R	G	I
I3	I	I	S	S	I	L	K	V	P	S	T	Q	G	I
I8	I	I	S	S	I	L	K	V	P	S	T	Q	S	I
I9	I	V	S	S	I	F	K	V	P	S	S	Q	S	I
I14	I	F	F	S	I	L	K	F	P	S	I	Q	D	I

F7	R	K
I15	H	K
I3	C	K
I8	H	K
I9	H	K
I14	Y	K

SUBSTITUTE SHEET

27/98

Figure 6D

					<u>V</u>														
F5	H	L	N	E	L	N	I	L	T	E	G	A	V	V					
I11	H	L	N	E	L	N	I	L	T	E	G	A	V	V					

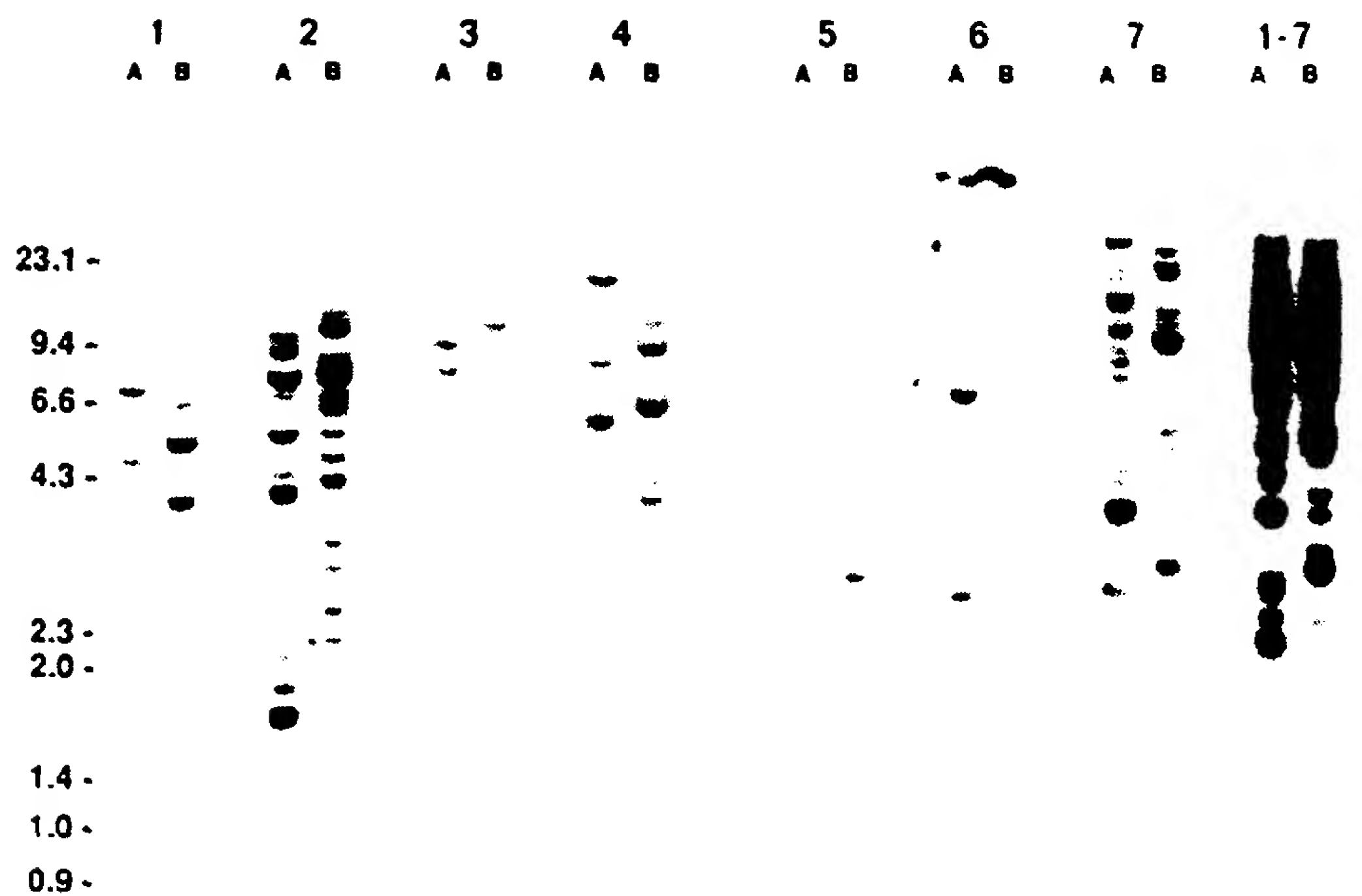
						<u>V</u>													
F5	N	V	T	P	F	V	C	I	L	I	S	Y	I	H					
I11	N	V	T	P	F	V	C	I	L	I	S	Y	I	H					

F5	I	T	C	A	V	L	R	V	S	S	P	R	G	G					
I11	I	T	W	A	V	L	R	V	S	S	P	R	G	G					

F5	W	K																	
I11	W	K																	

SUBSTITUTE SHEET

Figure 7

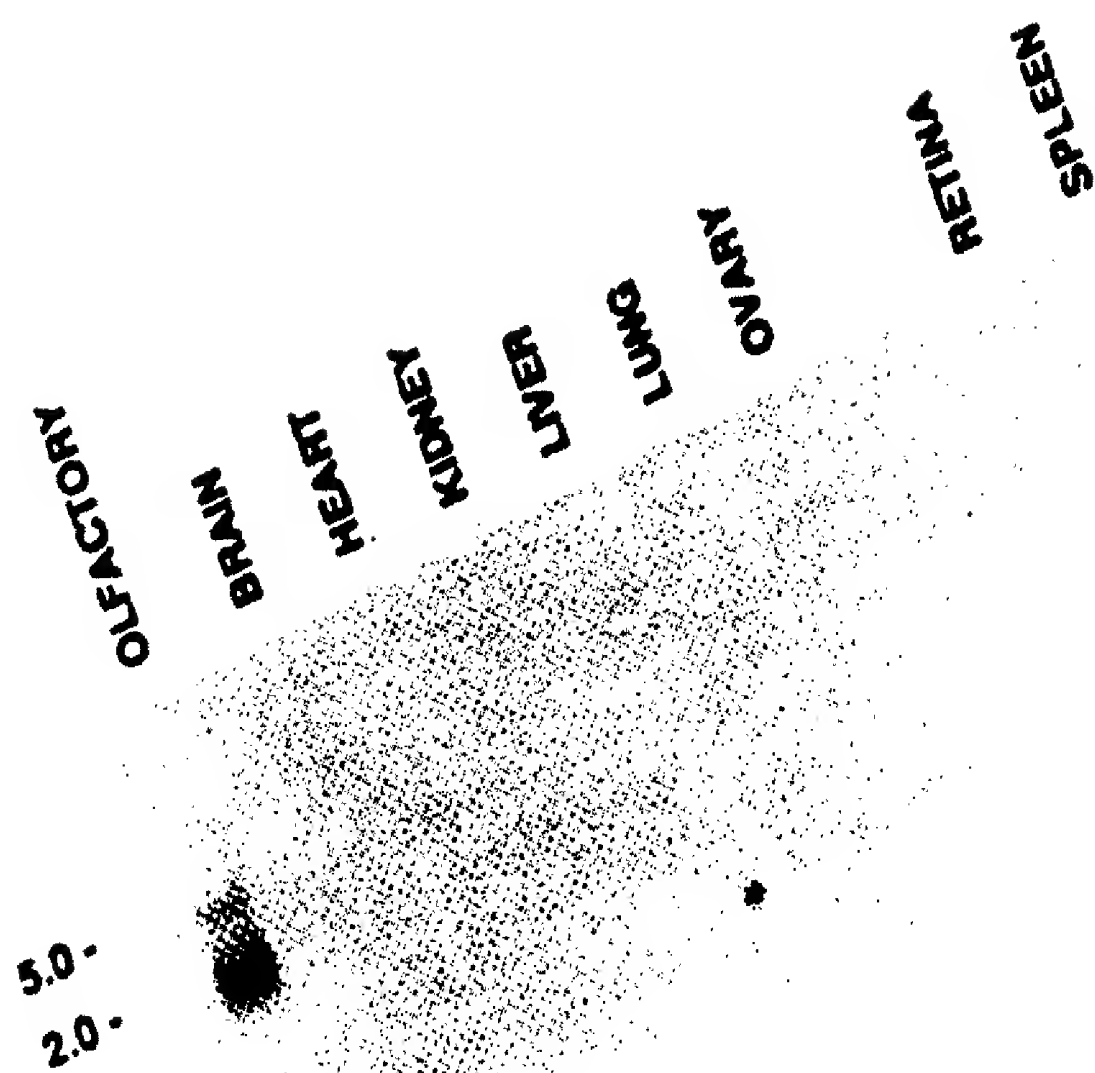


PCT/US92/02741

29/98

Figure 8

WO 92/17585



SUBSTITUTE SHEET

30/98

Figure 9A Translated sequence of F3T.D1S

10	20	30	40	50	60
* ATG GAC TCA AGC AAC AGG ACA AGA GTT TCA GAA TTT CTT CTT GGA TTT GTA GAA AAC	* M D S S N R T T R V S S E F L L L C F V E N	* 70	* 80	* 90	* 100
* AAA GAC CTA CAA CCC CTT ATT TAT GGT CTT TTT CTC TCT ATG TAC CTG GTT ACT GTC ATT	* K D L Q P L I Y G L F L S M Y L V T V I	* 130	* 140	* 150	* 160
* CGA AAC ATA TCC ATT ATT GTG GCT ATC ATT TCA GAT CCC TGT CTG CAC ACC CCC ATG TAT	* G N I S I I V A I I S D P C L H T P M Y	* 190	* 200	* 210	* 220
* TTC TTC CTC TCT AAC CTG TCC TTT GTG GAC ATC TGT TTC ATT TCA ACC ACT GTT CCA AAC	* F L S N L S F V D I C F I S T T V P K	* 250	* 260	* 270	* 280
* ATG TTA GTG AAC ATC CAG ACC CAA AAC AAT GTC ATC ACC TAT GCA GGA TGC ATT ACC CAG	* M L V N I Q T Q N Q N V I T Y A G C I T Q				

31/98

Figure 9B

310	*	320	*	330	*	340	*	350	*	360	*
ATA TAC TTT TTC TTG CTC TTT GTA GAA TTG GAC AAC TTC TTG CTG ACT ATC GCC TAT											
I Y F F L L F V E L D N F L L T I M A Y											
370	*	380	*	390	*	400	*	410	*	420	*
GAC CGT TAC GTA GCC ATC TGT CAC CCC ATG CAC TAC ACA GTT ATC ATG AAC TAC AAG CTC											
D R Y V A I C H P M H Y T V I M N Y K L											
430	*	440	*	450	*	460	*	470	*	480	*
TGT GGA TTT CTG GTT CTG GTA TCT TCG ATT GTA AGT GTT CTG CAT GCC TTG TTT CAA AGC											
C G F L V L V S I V S V L H A L F Q S											
490	*	500	*	510	*	520	*	530	*	540	*
TTG ATG ATG TTG GCG CTG CCC TTC TGC ACA CAT CTG GAA ATC CCA CAC TAC TTC TGT GAA											
L M M L A L P F C T H L E I P H Y F C E											
550	*	560	*	570	*	580	*	590	*	600	*
CCT AAT CAG GTG ATT CAA CTC ACC TGT TCT CAT CCA TTT CTT AAT GAT CTT GTG ATA TAT											
P N Q V I Q L T C S D A F L N D L V I Y											
610		620		630		640		650		660	

Figure 9C

TTT ACA CTT GTG CTG CTG GCT ACT GTT CCT CTT GCT GGC ATC TTC TAT TCT TAC TTC AAG	*	*	*	*	*	*	*
F T L V L L A T V P L A G I F Y S Y F K							
670	680	690	700	710	720		
ATA GTG TCC TCC ATA TGT GCT ATA TCG TCA GTT CAT GGG AAG TAC AAA GCA TTC TCC ACC	*	*	*	*	*		
I V S S I C A I S S S V H G K Y K A F S T							
730	740	750	760	770	780		
TGT GCA TCT CAC CTT TCA GTC GTG TCT TTA TTT TAC TGC ACA GGA CTA GGA GTG TAC CTC	*	*	*	*	*		
C A S H L S V V S L L F Y C T G L G V Y L							
790	800	810	820	830	840		
AGT TCT GCT GCA AAC AAC AGC TCA CAG GCA AGT GCC ACA GCC TCA GTC ATG TAC ACT GTA	*	*	*	*	*		
S A A N N S S Q A A S A T A S V M Y T V							
850	860	870	880	890	900		
GTT ACC CCT ATG GTG AAC CCT TTT ATC TAT AGT CTT AGG AAT AAA CAT GTT AAG AGT GTT	*	*	*	*	*		
PRONUC/TRA OPTION							
V T P M V N P F I Y S L R N K D V K S V							

SUBSTITUTE SHEET

Figure 9D

CTG AAA ACT CTT TGT GAG GAA GTT ATA AGG AGT CCA CCT TCC CTA CTT CAT TTC TTC	910	920	930	940	950	960
L K T L C E E V I R S P P S L L H F F	*	*	*	*	*	*
CTA GTG TTA TGT CAT CTC CCT TGT TTT ATT TTT TGT TAT TAA	970	980	990	1000		
L V L C H L P C F I F C Y -	*	*	*	*		

Translation begun with base no. 57

Translated to base no.1058

Sequence printed from base no. 57 to base no.1058

Sequence numbered beginning with base no. 57

Figure 10A Translated sequence of F5T.D1S

10	20	30	40	50	60
* ATG AGC AGC ACC AAC CAG TCC AGT GTC ACC GAG TTC CTC CTC GGA CTC TCC AGG CAG	* M S S T N Q S S V T E F L L L G L S R Q	* 70	80	90	100
* CCC CAG CAG CAG CTC CTC TTC CTC TTC CTC ATC ATG TAC CTG GCC ACT GTC CTG	* P Q Q Q L L F L L F L I M Y L A T V L	* 130	140	150	160
* GGA AAC CTG CTC ATC ATC CTG GCT ATT GGC ACA GAC TCC CGC CTG CAC ACC CCC ATG TAC	* G N L L I I L A I G T D S R L H T P M Y	* 190	200	210	220
* TTC TTC CTC AGT AAC CTG TCC TTT GTG GAT GTC TGC TTC TCC TCT ACC ACT GTC CCT AAA	* F L S N L S F V D V C F S S S T T V P K	* 250	260	270	280
* GTT CTG GCC AAC CAT ATA CTT GGG AGT CAG GCC ATT TCC TTC TCT GGG TGT CTC ACC CAG	* V L A N H I L G S Q A I S F S G C L T Q				

Figure 10B

CTG TAT TTT CTC GCT GTG TTT GGT AAC ATG GAC AAT TTC CTG CTG GCT CTG ATG TCC TAT L Y F L A V V F F G N M D N F L L A V M S Y	310	320	330	340	350	360
	*	*	*	*	*	*
GAC CGA TTT GTG GCC ATA TGC CAC CCT TTA CAC TAC ACA AAG ATG ACC CGT CAG CTC D R F V A I C H P L H Y T T K M T R Q L	370	380	390	400	410	420
	*	*	*	*	*	*
IGT GTC CTG CTT GTT GTG GGG TCA TGG GTT GTA GCC AAC ATG AAT TGT CTG TTG CAC ATA C V L L V V G S W V V A N M N C L L H I	430	440	450	460	470	480
	*	*	*	*	*	*
CTG CTC ATG GCT CGA CTC TCC TTC TGT GCA GAC AAC ATG ATC CCC CAC TTC TTC TGT GAT L L M A R L S F C A D N M I P H F C D	490	500	510	520	530	540
	*	*	*	*	*	*
GGA ACT CCC CTC CTG AAA CTC TCC TCC TCA GAC ACA CAT CTC AAT GAG CTG ATG ATT CTT G T P L L K L S C S D T H L N E L M I L	550	560	570	580	590	600
	*	*	*	*	*	*
	610	620	630	640	650	660

35 / 98

Figure 10C

ACA	GAG	GGA	GCT	GTG	GTC	ATG	GTC	ACC	CCA	TTT	GTC	TGC	ATC	CTC	ATC	TCC	TAC	ATC	CAC	*
T	E	G	A	V	V	M	V	T	P	F	V	C	I	L	I	S	Y	I	H	*
670						680			690			700				710			720	
ATC	ACC	TGT	GCT	GTC	CTC	AGA	GTC	TCA	TCC	CCC	AGG	GGA	GGA	TGG	AAA	TCC	TTC	TCC	ACC	*
I	T	C	A	V	L	R	V	S	S	P	R	G	G	W	K	S	F	S	T	*
730						740			750			760				770			780	
TGT	GGC	TCC	CAC	CTG	GCT	GTG	GTC	TGC	CTC	TTC	TAT	GGC	ACC	GTC	ATC	GCT	GTG	TAT	TTC	*
C	G	S	H	L	A	V	V	C	L	F	Y	G	T	V	I	A	V	Y	F	*
790						800			810			820				830			840	
AAC	CCA	TCA	TCC	TCT	CAC	TTA	GCT	GGG	AGG	GAC	ATG	GCA	GCT	GCA	GTG	ATG	TAT	GCA	GTG	*
PRONUC/TRA						OPTION														
N	P	S	S	S	H	L	A	C	R	D	M	A	A	A	V	M	Y	A	V	

Figure 10D

850	*	GTG ACC CCA ATG CTG AAC CCT TTC ATC TAT AGC CTG AGG AAC AGC GAC ATG AAA GCA GCT	880	*	900
		V T P M L N P F I Y S L R N S D M K A A			
910	*	TTA AGG AAA GTG CTC GCC ATG AGA TTT CCA TCT AAG CAG TAA	940	*	
		L R K V L A M R F P S K Q -			

Translation begun with base no. 62

Translated to base no.1003

; Sequence printed from base no. 62 to base no.1003

Sequence numbered beginning with base no. 62

Figure 11A Translated sequence of F6T.D1S

10	20	30	40	50	60
* ATG GCT TCG AGT ACT GGC CAG AAC CTG TCC ACA CCA GGA CCA TTC ATC TTG CTG GCC TTC M A W S T G Q N L S T P G P F I L L G F	* 70	* 80	* 90	* 100	* 110
* CCA GGG CCA AGG AGC ATG CGC ATT GGG CTC TTC CTG CTT TTC CTG GTC ATG TAT CTG CTT P G P R S M R I G L F L L F L V M Y L L	* 120	* 130	* 140	* 150	* 160
* ACG GTA GTT GGA AAC CTA GCC ATC ATC TCC CTG GTA GGT GCC CAC AGA TGC CTA CAG ACA T V V G N L A I I S L V G A H R C L Q T	* 170	* 180	* 190	* 200	* 210
* CCC ATG TAC TTC TTC CTC TGC AAC CTC TCC TTC CTG GAG ATC TGG TTC ACC ACA GCC TGC P M Y F F L C N L S F L E I W F T A C	* 220	* 230	* 240	* 250	* 260
* GTA CCC AAG ACC CTG GCC ACA TTT GCG CCT CGG GGT GGA GTC ATT TCC TTG GCT GCC TGT V P K T L A T F A P R G G V I S L A G C	* 270	* 280	* 290	* 300	* 310

Figure 11B

310	*	320	*	330	*	340	*	350	*	360	*
GCC ACA CAG ATG TAC TTT TCT TTG GGC TGT ACC GAG TAC TTC CTG GCT CTG											
A T Q M Y F V F S L G C T E Y F L A V											
370	*	380	*	390	*	400	*	410	*	420	*
ATG GCT TAT GAC CGC TAC CTG GGC ATC TGC CTG CCA CTG CGC TAT GGT GGC ATC ATG ACT											
M A Y D R Y L A I C L P L R Y G C I M T											
430	*	440	*	450	*	460	*	470	*	480	*
CCT GCG CTG GCG ATG CGG TTG GGC CTG GGA TCC TGG CTG TCT GGT TTT TCT GCA ATC ACA											
P G L A M R L L A A L G S W L C G F S A I T											
490	*	500	*	510	*	520	*	530	*	540	*
GTT CCT GCT ACC CTC ATT GCC CGC CTC TCT TTC TGT GGC TCA CGT GTC ATC AAC CAC TTC											
V P A T L I A R L S F C C S R V I N H F											
550	*	560	*	570	*	580	*	590	*	600	*
TTC TGT GAC ATT TCG CCC TGG ATA GTG CTT TCC TGC ACC GAC ACG CAG CTC GTG GAA CTC											
F C D I S P W I V L S C T D T Q V E L											
610		620		630		640		650		660	

SUBSTITUTE SHEET

Figure 11C

GTG TCC TTT GGC ATT GCC TTC TGT GTT ATT CTG GGC TCG TGT GGT ATC ACA CTA GTC TCC	*	*	*	*	*	*
V S F G I A F C V I L G S C G I T L V S						
670	680	690	700	710	720	
TAT GCT TAC ATC ATC ACT ACC ATC ATC AAG ATT CCC TCT GGC CGG GGC CAC CGC GCC	*	*	*	*	*	
Y A Y I I T T I I I K I P S A R G R H R A						
730	740	750	760	770	780	
TTC TCA ACC TGC TCA TCC CAT CTC ACT CTC CTC ATT TGG TAT GGC TCC ACC ATC TTC	*	*	*	*	*	
F S T C S S H L T V V L I W Y G S T I F						
790	800	810	820	830	840	
TTG CAT GTG AGG ACC TCG GTA GAG AGC TCC TTG GAC CTC ACC AAA GCT ATC ACA GTC CTC	*	*	*	*	*	
PRONUC/TRA	OPTION					
L H V R T S V E S S L D L T K A I T V L						

41 / 98

Figur 11D

850	*	AAC ACC ATT GTC ACA CCT GTG CTG AAC CCT TTC ATA TAT ACT CTG AGG AAC AAG GAT CTC	880	*	900	*
		N T I V T P V L N P F I Y T L R N K D V				
860	*		870	*		
910	*	AAG GAA GCT CTG CGC AGG ACG GTG AAG GGG AAG TGA	920	*	930	*
		K E A L R R R T V K G K -				

Translation begun with base no. 75

Translated to base no.1010

Sequence printed from base no. 75 to base no.1010

Sequence numbered beginning with base no. 75

SUBSTITUTE SHEET

Figure 12A Translated sequence of F12T.D1S

10	20	30	40	50	60
* ATG GAA TCA GGG AAC AGC ACA AGA AGA TTT TCA AGT TTT TTT CTT CTT GGA TTT ACA GAA	* M E S G N S T R R F S S F F L L G F T E	* 70	* 80	* 90	* 100
* AAC CCA CAA CTT CAC TTC CTC ATT TTT GCA CTA TTC CTG TCC ATG TAC CTG GTA ACA GTG	* N P Q L L H F L I F A L F L S M Y L V T V	* 110	* 120	* 130	* 140
* CTT GGG AAC CTG CTT ATC ATT ATG GCC ATC ATC ACA CAG TCT CAT TTG CAT ACA CCC ATG	* L G N L L I I M A I I T Q S H L H T P M	* 150	* 160	* 170	* 180
* 190	* 200	* 210	* 220	* 230	* 240
* TAC TTT TTC CTT GCT AAC CTA TCC TTT GTG GAC ATC TGT TTC ACC TCC ACC ACC ATC CCA	* Y F F L A N L S F V D I C F T S T T I P	* 250	* 260	* 270	* 280
* 290	* 300	* 310	* 320	* 330	* 340

Figure 12B

AAG ATG TTG GTA AAT ATA TAC ACC CAG AGC AAG AGC ATC ACC TAT GAA GAC TGT ATT AGC
K M L V N I Y T Q S K S I T Y E D C I S
310 320 330 340 350 360
* * * * *
CAG ATG TGT CTC TTC TTG GTT TTC GCA GAA TTG GGC AAC TTT CTC CTG GCT GTG ATG GCC
Q M C V F L V F A E L G N F L L A V M A
370 380 390 400 410 420
* * * * *
TAT GAC CGA TAT GTG GCT A-C TGT CAC CCA CTC TGT TAC ACA GTC ATT GTG AAC CAC CCG
Y D R Y V A X C H P L C Y T V I V N H R
430 440 450 460 470 480
* * * * *
CTC TGT ATC CTG CTT CTG CTG TCC TGG GTT ATC AGC ATT TTC CAT GCC TTC ATA CAG
L C I L L L L L S W V I S I F H A F I Q
490 500 510 520 530 540
* * * * *
AGC TTA ATT CTG CTA CAG TTG ACC TTC TGT GGA CAT GTG AAA ATC CCT CAC TTC TTC TGT
S L I V L Q L L T F C C D V K I P H F C
550 560 570 580 590 600
* * * * *
GAA CTT AAT CAG CTG TCC CAA CTC ACC TGT TCA GAC AAC TTT CCA AGT CAC CTC ATA ATG
E L N Q L S Q Q L T C S D N F P S H L I M

SUBSTITUTE SHEET

Figure 12C

AAT CTT GTA CCT GTT ATG TTG GCA GCC ATT TCC TTC AGT GGC ATC CTT TAC TCT TAT TTC	610	620	630	640	650	660
N L V P V M L A A I S F S G I L Y S F	*	*	*	*	*	*
AAG ATA GTA TCC TCC ATA CAT TCT ATC TCC ACA GTT CAG CGG AAG TAC AAG GCA TTT TCT	670	680	690	700	710	720
K I V S S I H S I S T V Q G K Y K A F S	*	*	*	*	*	*
ACT TGT GCC TCT CAC CTT TCC ATT GTC TCC TTA TTT TAT AGT ACA GGC CTC GGA GTG TAC	730	740	750	760	770	780
T C A S H L S I V S L F Y S T G L G V Y	*	*	*	*	*	*
GTC AGT TCT GCT GTG GTC CAA AGC TCA CAT TCT GCT GCA AGT GCT TCG GTC ATG TAT ACT	790	800	810	820	830	840
PRONUC/TRA OPTION	*	*	*	*	*	*
V S S A V V Q S S H S A A S A S V M Y T						

Figure 12D

850	860	870	880	890	900
* CTG GTC ACC CCC ATG CTG AAC CCC TTC ATT TAT AGT CTA AGG AAT AAA GAT GTG AAC AGA	* V V T P M L N P F I Y S L R N K D V K R	* 910	* 920	* 930	* 940
* GCT CTG GAA AGA CTG TTA GAA GGA AAC TGT AAA GTG CAT CAT TGG ACT GGA TGA	* A L E R L L E G N C K V H H T C -				

SUBSTITUTE SHEET

Translation begun with base no. 173

Translated to base no. 1126

Sequenc printed from base no. 173 to base no. 1126

Sequence numbered beginning with base no. 173

46 / 98

Figure 13A Translated sequence of I3T.D1S

10	20	30	40	50	60
* ATG AAC AAT CAA ACT TTC ATC ACC CAA TTC CTT CTC CTG GGA CTG CCC ATC CCT GAA GAA M N Q T F I T Q F L L L G L P I P E E	* 70	* 80	* 90	* 100	* 110
* CAT CAG CAC CTG TTC TAT GCC TTG TTC CTG GTC ATG TAC CTC ACC ACC ATC TTG GGA AAC H Q H L F Y A L F L V M Y L T T I L G N	* 130	* 140	* 150	* 160	* 170
* TTG CTA ATC ATT GTA CTT GTT CAA CTG GAC TCC CAG CTC CAC ACA CCT ATG TAT TTG TTT L L I I V L V Q L D S Q L H T P M Y L F	* 190	* 200	* 210	* 220	* 230
* CTC AGC AAT TTG TCT TTC TCT GAT CTA TGT TTT TCC TCT GTC ACA ATG CCC AAG CTG CTC L S N L S F S D L C F S S V T M P K L L	* 250	* 260	* 270	* 280	* 290
* CAG AAC ATG AGG AGC CAG GAC ACA TCC ATT CCC TAT GGA GGC TGC CTG GCA CAA ACA TAC Q N M R S Q D T S I P Y G C L A Q T Y					* 300

310	320	330	340	350	360
* TTC TTT ATG GTT TTT GGA GAT ATG GAG AGT TTC CTT CTT GTG GCC ATG GCC TAT GAC CGC	* F F M V F F G D M E S F L L V A M A Y D R	* 330	* 340	* 350	* 360
370	380	390	400	410	420
* TAT CTG GCC ATG TGC TTC CCT CTG CAT TAC ACC AGC ATC ATG AGC CCC AAG CTC TGT ACT	* Y V A I C F P L H Y T S I M S P K L C T	* 390	* 400	* 410	* 420
430	440	450	460	470	480
* TGT CTA CTG CTG TTA TTG TGG ATG CTG ACC ACA TCC CAT GCC ATG ATG CAC ACA CTG CTT	* C L V L L L W M L T T S H A M M H T L L	* 450	* 460	* 470	* 480
490	500	510	520	530	540
* GCA GCA AGA TTG TCT TTT TGT GAG AAC AAT CTG CTC AAC TTC TTC TTC TGT GAC CTA TTT	* A A R L L S F C E N N V V L N F F C D L F	* 510	* 520	* 530	* 540
550	560	570	580	590	600
* GTT CTC CTA AAG CTG GCC TGC TCA GAC ACT TAT ATT AAT GAG TTG ATG ATA TTT ATC ATG	* V L L K L A C S D T Y I N E L M I F I M	* 570	* 580	* 590	* 600
610	620	630	640	650	660

48 / 98

Figur 13C

AGT ACA CTC CTC ATT ATT ATT CCA TTC TTC CTC ATT GTT ATG TCC TAT CCA AGG ATC ATA	*	*	*	*	*	*
S T L L I I I I P F F F L I V M S Y A R I I						
670	680	690	700	710	720	
TCC TCT ATT CTT AAG GTT CCA TCT ACC CAA GGC ATC TGC AAG GTC TTC TCT ACC TGT GGT	*	*	*	*	*	*
S I L L K V P S T Q G I C K V F S T C G						
730	740	750	760	770	780	
TCC CAT CTG TCT GTA GTA TCA CTG TTC TAT GGG ACA ATT ATT GGT CTC TAC TTA TGT CCA	*	*	*	*	*	*
S H L S V V S S L F Y G T I I I G L Y L C P						

Figure 14A Translated sequence of I7T.D1S

10	20	30	40	50	60
* ATG GAG CGA AGG AAC CAC AGT GGG AGA GTG AGT GAA TTT GTG TTG CTG GGT TTC CCA GCT M E R R N H S G R V S E F V L L G F P A	* 20	* 30	* 40	* 50	* 60
70	80	90	100	110	120
* CCT GCC CCA CTG CGA GTA CTA TTT TTC CTT TCT TCT CTT CTG G-C TAT GTG TTG GTG TTC P A P L R V L L F F L S L L L X Y V L V L	* 80	* 90	* 100	* 110	* 120
130	140	150	160	170	180
* ACT GAA AAC ATG CTC ATC ATT ATA GCA ATT AGG AAC CAC CCA ACC CTC CAC AAA CCC ATG T E N M L I I I A I R N H P T L H K P M	* 140	* 150	* 160	* 170	* 180
190	200	210	220	230	240
* TAT TTT TTC TTG GCT AAT ATG TCA TTT CTG CAG ATT TCG TAT GTG ACT GTT ACG ATT CCT Y F F L A N M S F L E I W Y V T V T I P	* 200	* 210	* 220	* 230	* 240
250	260	270	280	290	300
* AAG ATG CTC GCT GCC TTC ATT GGT TCC AAG GAG AAC CAT GGA CAG CTG ATC TCC TTT GAG K M L A G F I G S K E N H G Q L I S F E	* 260	* 270	* 280	* 290	* 300

SUBSTITUTE SHEET

Figure 14B

310	*	320	*	330	*	340	*	350	*	360	*
GCA TGC ATG ACA CAA CTC TAC TTT TTC CTC GGC TTG GGT TGC ACA GAG TGT GTC CTT CTT											
A C M T Q L Y F F L G L G C C T E C V L L											
370	*	380	*	390	*	400	*	410	*	420	*
GCT GTG ATG GCC TAT GAC CGC TAT GTG GCT ATC TGT CAT CCA CTC CAC TAC CCC GTC ATT											
A V M A Y D R Y V A I C H P L H Y P V I											
430	*	440	*	450	*	460	*	470	*	480	*
GTC AGT AGC CGG CTA TGT GTG CAG ATG GCA GCT GGA TCC TCG GCT GGA GGT TTT GGT ATC											
V S S R L C V Q M A A G S W A G G F G I											
490	*	500	*	510	*	520	*	530	*	540	*
TCC ATG GTT AAA GTT TTC CTT ATT TCT CGC CTG TCT TAC TGT GGC CCC AAC ACC ATC AAC											
S M V K V F L I S R L S Y C G P N T I N											
550	*	560	*	570	*	580	*	590	*	600	*
CAC TTT TTC TGT GAT GTG TCT CCA TTG CTC AAC CTG TCA TGC ACT GAC ATG TCC ACA GCA											
H F F C D V S P L L N L S C T T D M S T A											

Figure 14C

610	*	620	*	630	*	640	*	650	*	660	*
GAG CTT ACA GAC TTT GTC CTG GCC ATT TTT ATT CTG CTG GGA CCG CTC TCT GTC ACT GCG											
E L T D F V L A I F I L L G P L S V T G											
670	*	680	*	690	*	700	*	710	*	720	*
GCA TCC TAC ATG GCC ATC ACA GGT GCT GTG ATG CGC ATC CCC TCA GCT GCT GCG CGC CAT											
A S Y M A I T G A V M R I P S A A G R H											
730	*	740	*	750	*	760	*	770	*	780	*
AAA GCC TTT TCA ACC TGT GCC TCC CAC CTC ACT GTT GTG ATC ATC TTC TAT GCA GCC ACT											
K A F S T C A S H L T V V I I F Y A A S											
790	*	800	*	810	*	820	*	830	*	840	*
ATT TTC ATC TAT GCC AGG CCT AAG GCA CTC TCA GCT TTT GAC ACC AAC AAG CTG CTC TCT											
I F I Y A R P K A L S A F D T N K L V S											
850	*	860	*	870	*	880	*	890	*	900	*
GTA CTC TAC GCT ATT GTA CCG TTG TTC AAT CCC ATC ATC TAC TGC TTG CGC AAC CAA											
PRONUC/TRA											
V L Y A V I V P L F N P I I Y C L R N Q											

Figure 14D

910	*	920	*	930	*	940	*	950	*	960	*
GAT GTC AAA AGA GCG CTA CGT CGC ACG CTG CAC CTG GCC CAG GAC CAG GCG AAT ACC											
D V K R A L R R T L H L A Q D Q E A N T											
970	*	980	*								
AAC AAA GCG AGC AAA ATT GGT TAG											
N K G S K I G -											

Translation begun with base no. 119

Translated to base no. 1102

Sequence printed from base no. 119 to base no. 1102

Sequence numbered beginning with base no. 119

SUBSTITUTE SHEET

53/98

Figure 15A Translated sequence of I8T.D1S

10	20	30	40	50	60
* ATG AAC AAC AAA ACT GTC ATC ACC CAT TTC CTC CTC CTG GGA TTG CCC ATC CCC CCA GAG M N K T V I T H F L L L G L P I P P E	* 70	* 80	* 90	* 100	* 110
* CAC CAG CAA CTG TTC TTT GCC CTG TTC CTC ATC ATG TAC CTC ACC ACC TTT CTG GGA AAC H Q Q L F F A L L F L I M Y L T T F L G N	* 130	* 140	* 150	* 160	* 170
* CTG CTA ATT GTT GTC CTT GTT CAA CTG GAC TCT CAT CTC CAC ACA CCC ATG TAC TTG TTT L L I V V L V Q L D S H L H T P M Y L F	* 190	* 200	* 210	* 220	* 230
* CTC AGC AAC TTG TCC TTC TCT GAT CTC TGC TTT TCC TCT GTT ACA ATG CTG AAA TTG CTC L S N L S F S D L L C F S S V T M L K L L	* 250	* 260	* 270	* 280	* 290
* CAA AAT ATA CAG AGC CAA GTA CCA TCT ATA TCC TAT GCA GGA TGC CTG ACA CAG ATA TTC Q N I Q S Q V P S I S Y A G C L T Q I F					

Figur 15B

TTC TTT TTG TTG TTT GGC TAC CTT GGG AAT TTC CTT CTT GTA GCC ATG GCC TAT GAC CGC	310	320	330	340	350	360
F L L L F G Y L L G N F L L V A M A Y D R	*	*	*	*	*	*
TAT CTG GCC ATC TGC TTC CCT CTG CAT TAT ACC AAC ATC ATG AGC CAT AAG CTC TGT ACT	370	380	390	400	410	420
Y V A I C F P L H Y T N I M S H K L C T	*	*	*	*	*	*
IGT CTC CTG GTA TTT TGG ATA ATG ACA TCA TCT CAT GCC ATG ATG CAC ACC CTG CTT	430	440	450	460	470	480
C L L L V F W I M T S S H A M M H T L L	*	*	*	*	*	*
GCA GCA AGA TTG TCT TTT TGT GAG AAC AAT GTA CTC CTC AAC TTT TTC TGT GAC CTG TTT	490	500	510	520	530	540
A A R L L S F C E N N V L L L N F F C D L F	*	*	*	*	*	*
GTT CTC CTA AAG TTG GCC TGC TCA GAC ACT TAT GTT AAT GAG TTG ATG ATA CAT ATC ATG	550	560	570	580	590	600
V L L K L A C S D T Y V N E L M I H I M	*	*	*	*	*	*
	610	620	630	640	650	660

GGC	GTG	ATC	ATC	ATT	GTT	ATT	CCA	TTC	GTG	CTC	ATT	GTT	ATA	TCC	TAT	GCC	AAG	ATC	ATC	*	*	*
G	V	I	I	I	V	I	P	F	V	L	I	V	I	S	Y	A	K	I	I			
670						680			690				700			710			720			
TCC	TCC	ATT	CTT	AAG	GTT	CCA	TCT	ACT	CAA	AGC	ATT	CAC	AAG	GTC	TTC	TCC	ACT	TGT	GGT	*	*	*
S	S	I	L	K	V	P	S	T	Q	S	I	H	K	V	F	S	T	C	G			
730						740			750				760			770			780			
TCT	CAT	CTC	TCT	GTG	GTG	TCT	CTG	TTC	TAC	GGC	ACA	ATT	ATT	GGT	CTC	TAT	TTA	TGT	CCA	*	*	*
S	H	L	S	V	V	S	L	F	Y	G	T	I	I	G	L	Y	L	C	P			
790						800			810				820			830			840			
TCA	GGT	GAT	AAT	TTT	AGT	CTA	AAG	GGG	TCT	GCC	ATG	GCT	ATG	ATG	TAC	ACA	GTG	GTA	ACT	*	*	*
PRONUC/TRA OPTION																						
S	G	D	N	F	S	L	K	G	S	A	M	A	M	M	Y	T	V	V	T			
850						860			870				880			890			900			
CCA	ATG	CTG	AAC	CCG	TTC	ATC	TAC	AGC	CTA	AGA	AAC	AGA	GAC	ATG	AAG	CAG	GCC	CTA	ATA	*	*	*
P	M	L	N	P	F	I	Y	S	L	R	N	R	D	M	K	Q	A	L	I			

56/98

Figure 15D

	910		920		930		9
	*		*		*		
AGA GTT ACC TGT AGC AAG AAA ATC TCT CTG CCA TGG TAG							
R V T C S K K I S L P W -							

Translation begun with base no. 57
 Translated to base no. 995
 Sequence printed from base no. 57 to base no. 995
 Sequence numbered beginning with base no. 57

SUBSTITUTE SHEET

Figure 16A Translated sequence of I9T.D1S

	10	20	30	40	50	60
ATC ACT AGA AGA AAC CAA ACT GCC ATC TCT CAG TTC TTC CTT TTC CTG GGC CTC CCA TTC CCC	* *	* *	* *	* *	* *	* *
M T R R N Q T A I S Q F F L L G L P F P						
CCA GAG TAC CAA CAC CTG TTC TAT GCC CTG TTC CTG GGC ATG ATG TAC CTC ACC ACT CTC CTC	70 *	80 *	90 *	100 *	110 *	120 *
P E Y Q H L F Y A L F L A M Y L T T L L						
GGG AAC CTC ATC ATC ATC CTC ATT CTA CTG GAC TCC CAT CTC CAC ACA CCC ATG TAC	130 *	140 *	150 *	160 *	170 *	180 *
G N L I I I I L I L L D S H L L H T P M Y						
TTC TTT CTC AGC AAT TTA TCC TTT GCC GAC CTC TGT TTT TCC TCT GTC ACA ATG CCC AAG	190 *	200 *	210 *	220 *	230 *	240 *
L F L S N L S L S F A D L C F S S V T M P K						
	250	260	270	280	290	300

Figure 16B

TTG TTG CAG AAC ATG CAG AGC CAA GTT CCA TCC ATC CCC TAT GCA GGG TGC CTG GCA CAG	*	*	*	*	*	*	*	*
L L Q N M Q S Q V P S I P Y A G C L A Q								
310	*	320	*	330	*	340	*	350
ATA TAC TTC TTT CTG TTT TTT GGA GAC CTT GGA AAC TTC CTG CTT CTG CCC ATG GCC TAT								
I Y F F L F F F G D L L G N F L L V A M A Y								
370	*	380	*	390	*	400	*	410
GAC CGC TAT GTG GCC ATC TGC TGC CCC CTT CAT TAC ATG AGC ATC ATG AGC CCC AAG CTC								
D R Y V A I C F P L L H Y M S I M S P K L								
430	*	440	*	450	*	460	*	470
TGT GTG AGT CTG CTG CTG TCC TGG CTG CTG ACT ACC TTC CAT CCC ATG CTG CAC ACC								
C V S L V V L S W V L T T F H A M L H T								
490	*	500	*	510	*	520	*	530
CTG CTC ATG GCC AGA TTG TCA TTC TGT GAG GAC AGT GTG ATC CCT CAC TAT TTC TGT CAT								
L L M A R L S F C E D S V I P H Y F C D								
550	*	560	*	570	*	580	*	590
ATG TCT ACT CTG CTG AAA GTG GCT TGT TCT GAC ACC CAT GAT AAT GAA TTA GCA ATA TTT								
M S T L L K V A C S D T H D N E L A I F								

SUBSTITUTE SHEET

SUBSTITUTION QUEST

BNSDOCID: <WO 9217585A1 1A>

Figur 16D

850	*	860	*	870	*	880	*	890	*	900	*
GTG ACA CCC ATG CTG AAC CCC TTC ATC TAC AGC CTA AGA AAC AGA GAC ATA AAA GAT GCA											
V T P M L N P F I Y S L R N R D I K D A											
910	*	920	*	930	*	940	*				
TTA GAA AAA ATA ATG TGC AAA AAG CAA ATT CCC TCC TTT CTA TGA											
L E K I M C K K Q I P S F L -											

Translation begun with base no. 200

Translated to base no.1144

Sequence printed from base no. 200 to base no.1144

Sequence numbered beginning with base no. 200

SUBSTITUTE SHEET

61 / 98

Figure 17A Translated sequence of I14T.D1S

10	20	30	40	50	60
* ATG ACT GGA AAT AAC CAA ACT TTG ATC TTC GAG TTC CTC CTC GGT CTC CCC ATC CCA M T G N N Q T L I L E F L L G L P I P	* 70	* 80	* 90	* 100	* 110
* TCA GAG TAT CAT CTC CTG TTC TAT GCC CTG TTC CTC GGC ATG TAC CTC ACC ATC ATC CTG S E Y H L L F Y A L F L A M Y L T I I L	* 130	* 140	* 150	* 160	* 170
* GGA AAC CTG CTA ATC ATT GTC CTT GTC GAC TCT CAT CTC CAC ATG CCC ATG TAC G N L L I I V L R L D S H L H M P M Y	* 190	* 200	* 210	* 220	* 230
* TTG TTT CTC AGC AAC TTG TCC TTC TCT GAC CTC TGC TGC TTT TCC TCT CTC ACA ATG CCC AAA L F L S N L S F S D L C F S S V T M P K	* 250	* 260	* 270	* 280	* 290
* TTG CTT CAG AAC ATG CAG AGC CAA GTA CCA TCT ATA TCC TAT ACA CGC TGC CTG ACA CAG L L Q N M Q S Q V P S I S Y T G C L T Q					

SUBSTITUTE SHEET

62 / 98

Figure 17B

CTG TAC TTC TTT ATG GTT TTT GGA GAT ATG GAG AGC TTC CTT CTT GTG CTC ATG GCC TAT	310	320	330	340	350	360
L Y F F M V F G D M E S F L L V V M A Y	*	*	*	*	*	*
GAC CGC TAT GTG GCC ATT TGC TTT CCT TTG CGT TAC ACC ACC ATC ATG AGC ACC AAG TTC	370	380	390	400	410	420
D R Y V A I C F P L R Y T T I M S T K F	*	*	*	*	*	*
TGT GCT TCA CTA GTG CTA CTT CTG TGG ATG CTG ACG ATG ACC CAT GCC CTG CTG CAT ACC	430	440	450	460	470	480
C A S L V L L L L W M L T M T H A L L H T	*	*	*	*	*	*
CTA CTC ATT GCT AGA TTG TCT TTT TGT GAG AAG AAT GTG ATT CTT CAC TTT TTC TGT GAC	490	500	510	520	530	540
L L I A R L L S F C E K N V I L L H F C D	*	*	*	*	*	*
ATT TCT GCT CTT CTG AAG TTG TCC TCA GAC ATT TAT GTT AAT GAG CTG ATG ATA TAT	550	560	570	580	590	600
I S A L L K L L S C S D I Y V N E L M I Y	*	*	*	*	*	*
	610	620	630	640	650	660

SUBSTITUTE SHEET

Figure 17C

ATC	TTC	GGT	GGA	CTC	ATC	ATT	ATT	ATC	CCA	TTC	CTA	TTA	ATT	GTT	ATG	TCC	TAT	GTT	AGA	*
I	L	G	G	L	I	I	I	I	P	F	L	L	I	V	M	S	Y	V	R	*
670																				
ATT	TTC	TTC	TCC	ATT	TTG	AAG	TTT	CCA	TCT	ATT	CAG	GAC	ATC	TAC	AAG	GTA	TTC	TCA	ACC	*
I	F	F	S	I	L	K	F	P	S	I	Q	D	I	Y	K	V	F	S	T	*
730																				
TGT	GGT	TCC	CAT	CTG	TCT	GTG	GTG	ACC	TTG	TTT	TAT	GGG	ACA	ATT	TTT	GCT	ATC	TAC	TTA	*
C	G	S	H	L	S	V	V	T	L	F	Y	G	T	I	F	G	I	Y	L	*
790																				
TGT	CCA	TCA	GGT	AAT	AAT	TCT	ACT	GTG	AAG	GAG	ATT	GCC	ATG	GCT	ATG	ATG	TAC	ACA	CTC	*
PRONUC/TRA OPTION																				
C	P	S	G	N	N	S	T	V	K	E	I	A	M	A	M	M	Y	T	V	*
850																				
GTG	ACT	CCC	ATG	CTG	AAT	CCC	TTT	ATC	TAC	AGC	CTG	AGG	AAC	AGA	CAC	ATG	AAA	AGG	GCC	*
V	T	P	M	L	N	P	F	I	Y	S	L	R	N	R	D	M	K	R	A	*
880																				
890																				
900																				

SUBSTITUTE SHEET

64/98

Figure 17D

	910		920	930	9
	*	*	*	*	
CTA ATA AGA GTT ATC TGC ACT AAG AAA ATC TCT CTG TAA					
L I R V I C T K K I S L -					

Translation begun with base no. 64
 Translated to base no.1002
 Sequence printed from base no. 64 to base no.1002
 Sequence numbered beginning with base no. 64

SUBSTITUTE SHEET

65/98

Figure 18A Translated sequence of 115T.D1S

10	20	30	40	50	60
* ATG ACA GAA GAG AAC CAA ACT GTG ATC TCC CAG TTC CTT CTC CTT TTC CTC CCC ATC CCC M T E E N Q T V I S Q F L L L F L P I P	* 70	* 80	* 90	* 100	* 110
* TCA GAG CAC CAC GTG TTC TAC GCC CTG TTC CTG TCC ATG TAC CTC ACC ACT GTC CTG S E H Q H V F Y A L F L S M Y L T T V L	* 130	* 140	* 150	* 160	* 170
* GGG AAC CTC ATC ATC ATC CTC ATT CAC CTG GAC TCC CAT CTC CAC ACA CCC ATG TAC G N L I I I I L I H L D S H L H T P M Y	* 190	* 200	* 210	* 220	* 230
* TTG TTT CTC AGC AAC TTG TCC TTC TCT GAT CTC TGC TGC TTT TCC TCT GTT ACG ATG CCC AAG L F L S N L S F S D L C F S S V T M P K	* 250	* 260	* 270	* 280	* 290
* TTG TTG CAG AAC ATC CAG AGC CAA GTT CCA TCC ATC CCC TTT GCA GCC TGC CTG ACA CAA					* 300

Figure 18B

TTA TAC TTT TAC CTG TAT TTT GCA GAC CTT GAG AGC TTC CTG CTT CTC GCC ATG GCC TAT	310	320	330	340	350	360
L Y F Y L Y Y F A D L L E S F L L V A M A Y	*	*	*	*	*	*
GAC CGC TAT GTG GCC ATC TGC TCC CTC GTG CTG ACC TAC ATG AGC ATC ATG AGC CCC AAG CTC	370	380	390	400	410	420
D R Y V A I C F P L L H Y M S I M S P K L	*	*	*	*	*	*
TGT GTG AGT CTG GTG GTG TCC TCG GTG CTG ACC ACC TTC CAT GCC ATG CTG CAC ACC	430	440	450	460	470	480
C V S L V V L S W V L T T F H A M L H T	*	*	*	*	*	*
CTG CTC ATG GCC AGA TTG TCA TTC TGT GCG GAC AAT ATG ATC CCC CAC TTT TTC TGT GAT	490	500	510	520	530	540
L L M A R L S F C A D N M I P H F C D	*	*	*	*	*	*
ATA TCT CCT TTA TTG AAA CTG TCC TCG TCT GAC ACC CAT GTT AAT GAG TTG CTC ATA TTT	550	560	570	580	590	600
I S P L L K L S C S D T H V N E L V I F	*	*	*	*	*	*
	610	620	630	640	650	660

Figure 25B

301 XXXXXXXXXXXXXXXXXXXXXXXXXX
-----+-----+-----+360

[illegible][illegible]

-H 3 S M S . L i L L L L L L L L L L L L L

ATTGTCGCATCCTCATCTCTTACATCTACATCACCACATGCAGTCCCTCAGAGTCTCATC
421 -----+-----+-----+-----+-----+-----+-----+480

P V C I L I S Y I Y I T N A V L R V S S -

CTTTAGGGGACGATGGAAAGCCCTTCTCCACCTGTGGCTACACCTGGCTGTGGTCTGCCT
481 -----+-----+-----+-----+-----+-----+-----+540

F R G G W K A F S T C G S H L A V V C L

514

[illegible]

88/99

Figure 26B

```
301  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX+360
    ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? -
361  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTGATCATGGTCACCCC+420
    ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? -
    ATTGTGTCATCCTCATCTTTACATCTACATCACCATGCAGTCTCAGAGTCTCATC+480
    F V C I L I S Y I Y I T N A V L R V S S -
    CTTTAGGGAGGATGGAAAGCCTTCTCCACCTGTGGCTCACACCTGGCTGTGGTCTGCCT+540
    F R G G W K A F S T C G S H L A V V C L -
    CTTCTATGGCACCATCATTTGCTGTGTATTTCATCTCTGTATCTTCCCATCTGAGAA+600
    F Y C T I I A V Y P N P V S S H S S E K -
    GGACACTGCAGCACTGTGCTATACACAGTGGTGGTCCCATGTTG
601  -----+----- 646
    D T A A T V L Y T V V T P M L -
```

SUBSTITUTE SHEET

89/99

Figure 27A

J15

```
1  TATCTGCAACCCCTCTGCGCTACCCAGTCTCATGAGCGCGCGGTGTCCTGCTCATGCT 60
   I C N P L R Y P V L M S Q R V C L L M V
61  CGTGGCCTCCTGGTTGGAGGATCCCTCAAGCCCTCCATTCTCTGACCCCTTCA
   V A S W L G G S L N A S I Q T S L T L Q -
121 GTTCCCCTACTGTGGATCAGGAAGATCTCCCACTTCTCTGAGGTGCCCTCGCTGCT
   F P Y C G S R R I S H F P C E V P S L L -
181 GAXXXTGGCCCTGTGCAGACACTGAAGCCTATGAGCAGGTACTATTGTGACAGCGGTGCT
   ? ? A C A D T B A Y E Q V L F V T G V V -
240
```

SUBSTITUTE SHEET

90/99

Figure 27B

GGTCCTCCTGGTGCCCATTAACATTACTGCTTATGCCCTCATCCTGGCTGCTGT
241 -----+-----+-----+-----+-----+-----+
V L L V P I T P I T A S Y A L I L A A V
GCTCCGAATGCACCTCTGGAGGGAGTCAGAGGCCCTAGCCACATGCTCTCTCACCT
301 -----+-----+-----+-----+-----+-----+360
L R M H S A E G S Q K A L A T C S S H L -
GACAGTCGTCAATCTTCTATGGGCCCTTGTCTACACCTACATGTTACCTGCTTCCTA
361 -----+-----+-----+-----+-----+-----+420
T V V N L F Y C P L V Y T Y M L P A S Y -
TCACTCACCGCCAGACGACATAGTATCCGTCTTTTACACCGTCTCACACCCATGCT
421 -----+-----+-----+-----+-----+-----+480
H S P C Q D D I V S V F Y T V L T P M L -

T
481 - 481
A

SUBSTITUTE SHEET

Figure 28A

J16

1 CATCTGTAGGCTCTTCACTATCCTACCCCTCATGACCCAGACACTGTGTGCCAAGATTGC + 60
I C R P L H Y P T L M T Q T L C A K I A -

61 CACTGGTTCCTGGTGGAGGCTTCGCTGGGCCAGTGGTAGAAATTTCCTGGTGTCTCG + 120
T G C W L G G L A G P V V B I S L V S R -

121 TCTCCTTTTGTGGCCCCAATCACATTCAACACATCTTTGTGATTTCACCCTGTGCT + 180
L L F C G P N H I Q H I F C D F P P V L -

181 GAGCTTGGCTTGTACTCATACATCAGTGAAATGTCCTGGTAGATTTTATTATAAACCCTCG + 240
S L A C T D T S V N V L V D F I I N L C -

241 CAAGATCCTGGCCACCTTCCTGCTGATCCTGAGCTCCTACTTGCAGATAATCCGCACAGT + 300
R I L A T F L L I L S S Y L Q I I R T V -

91/99

SUBSTITUTE SHEET

92/99

Figure 28B

181 GAGCTTGGCTTGTAATACATCAGTGAAATGCTCCTGGTAGATTTTATATAACCTCTC
S L A C T D T S V N V L V D P I I N L C - +240 -

241 CAAGATCCTGGCCACCTTCCTGCTGATCCTGAGCTCCTACTTGCAGATAATCCGCACAGT
K I L A T F L L I L S S Y L Q I I R T V - +300 -

301 GCTCAAGATTCCTTCAGCTGCAGGCAAGAAGCAATCTCGACTTGTGCTCCCATCT
L K I P S A A O K K A P S T C A S H L - +360 -

361 CACTGTGGTCTCATCTTCTATCGAGCATCCTTTCATGTATGTGCGCCTGANGAAGAC
T V V L I F Y G S I L P M Y V R L K K S - +420 -

421 TTACTCCCTTGACTACGACAGAGCCCTTGGCAGTAGTCTACTCCGTGTTACCCCTTTCCT
Y S L D Y D R A L A V V Y S V V T P F L - +480 -

G
481 - 481

SUBSTITUTE SHEET

93/99

Figure 29A

J17

```

1  AATCTCAACCCACTGCTTTATTCCACCAAAATGCCACACAAGTCTGTATCCAGTTGGT
   -----+-----+-----+-----+-----+-----+-----+-----+
60  I C N P L L Y S T K M S T Q V C I Q L V -
   -----+-----+-----+-----+-----+-----+-----+-----+

61  TGCAGGATCTTATAGGGGTTTCTTAATACTGCCCTCATGTTTACTTTTCTC
   -----+-----+-----+-----+-----+-----+-----+-----+
120  A G S Y I G C P L N T C L I M F Y F F S -
   -----+-----+-----+-----+-----+-----+-----+-----+

121  TTTTCTCTCTGTGGCCAAATATAGTTGATCATTTTTCGTGTCATTTTGCCTTCTTXXT
   -----+-----+-----+-----+-----+-----+-----+-----+
180  F L P C G P N I V D H F F C D F A P ? ? -
   -----+-----+-----+-----+-----+-----+-----+-----+

181  GGAACTTTCGTGCTCTGATGTGAGTGTCTCTGTAGTTGTATGTCATTTTCTGCTGGCTC
   -----+-----+-----+-----+-----+-----+-----+-----+
240  E L S C S D V S V S V V V M S F S A G S -
   -----+-----+-----+-----+-----+-----+-----+-----+

241  AGTTACTATGATCACAGTGTATTATCATAGCCATCTCTATTCTTACATCCTCATCACCAT
   -----+-----+-----+-----+-----+-----+-----+-----+
300  V T M I T V F I I A I S Y S Y I L I T I -
   -----+-----+-----+-----+-----+-----+-----+-----+

```

SUBSTITUTE SHEET

94/99

Figure 29B

301 CCTGAAGATGTCCTCACTGACGGCCGTCACAAGGCTTTCCTCCACATGTACCTCCACCT
-----+-----+-----+-----+-----+-----+360
L K M S S T E G R H K A F S T C T S H L -

361 CACTGCAGTCACTCTCTACTATGGCACCATTACCTTCATTATGTGATGCCCAAGTCCAC
-----+-----+-----+-----+-----+-----+420
T A V T L Y Y G T I T F I Y V M P K S T -

421 ATACTCTACAGACCAGAACAAGGTGGTGTCTGTCTTTACATGGTGGTGATCCCAATGTT
-----+-----+-----+-----+-----+-----+480
Y S T D Q N K V V S V F Y M V V I P M L -

G
481 - 481

SUBSTITUTE SHEET

95/99

Figure 30A

J19

1 TATCTGCCACCCCTCTGAAGTACACAGTTATCATGAATCACTATTTTGTGTGATGCTOCT
I C H P L K Y T V I M N H Y P C V M L L - 60
61 GCTCTTCTCTGTCTGTTAGCACTTCCACATTTTAATGGTGTGAT
L F S V F V S I A H A L F H I L M V L I - 120
121 ACTGACTTTCAGCACAAACTGAATCCCTCACTTTTCTGTGAGCTGGCTCATATCAT
L T F S T K T E I P H F F C E L A H I I - 180
181 CAAACTTACCTGTTCGGATAATTTATCAACTATCTGCTGATATACACAGAGTCTGTCTT
K L T C S D N F I N Y L L I Y T E S V L - 240
241 ATTTTGTGGTTCATATGTAGGATCATTTTGTCTTATATTACACTGTATCCTCAGT
F F G V H I V G I I L S Y I Y T V S S V - 300

SUBSTITUTE SHEET

96/99

Figure 30B

```
301  TTTAAGAAATGTCATTATTGGGAGGAATGTATAAAGCCTTTTCAACATGTGGATCTCATTT
    L R M S L L G L G M Y K A F S T C G S H L -
360
361  GTCGGTGTCTCTGTTTATGGCACAGCTTTTNGGGGTACACATAAGCTCTCCACTTACTG
    S V V S V L W H R F W G T H K L S T Y * -
420
421  ACTCTCCAAGGAAGACTGTAGTGGCTTCAGTGATGTACACTGTGGTTACTCAGATGCTG
    L S K E D C S G F S D V H C G Y S D A -
479
```

SUBSTITUTE SHEET

97/99

Figure 31A

1 AATCTGCTACCCACTGAGGTACCTTCTCATCATGAGCTGGGTGGTGCACAGCACTGTC
I C Y P L R Y L L I M S W V V C T A L S - 60
61 CGTGGCAATCTGGGTACATAGGCTTTGTGCTCCGTTATACCTCTCTCTCTCAGATCCT
V A I W V I G F C A S V I P L C F T I L - 120
121 CCCACTCTGTGGTCCTACGTGATTATCTTTCTCGGAGCTGCCCATCCTTCGCA
P L C G P Y V V D Y L F C E L P I L L H - 180
181 CCTGTCTGCACACATACATCTCTGCTGGAGXXXXXXXDDDDDDDDDDDDDDDD
L F C T D T S L L E ? ? ? ? ? ? ? ? ? ? - 240
241 XXXXXXXXXXXCCCTTCCTCTGATGTTCTCTCTACCTTCGCATCCGTGGCTGTG
? ? ? P F L L I V L S Y L R I L V A V - 300

J20

SUBSTITUTE SHEET

Figure 31B

301 ATAAGAAATAGACTCAGCTGAGGCGCAGAAAAGCCCTTTTCAACTTGCTGCACACTTG
-----+-----+-----+-----+-----+360
 I R I D S A E O R K K A F S T C A S H L

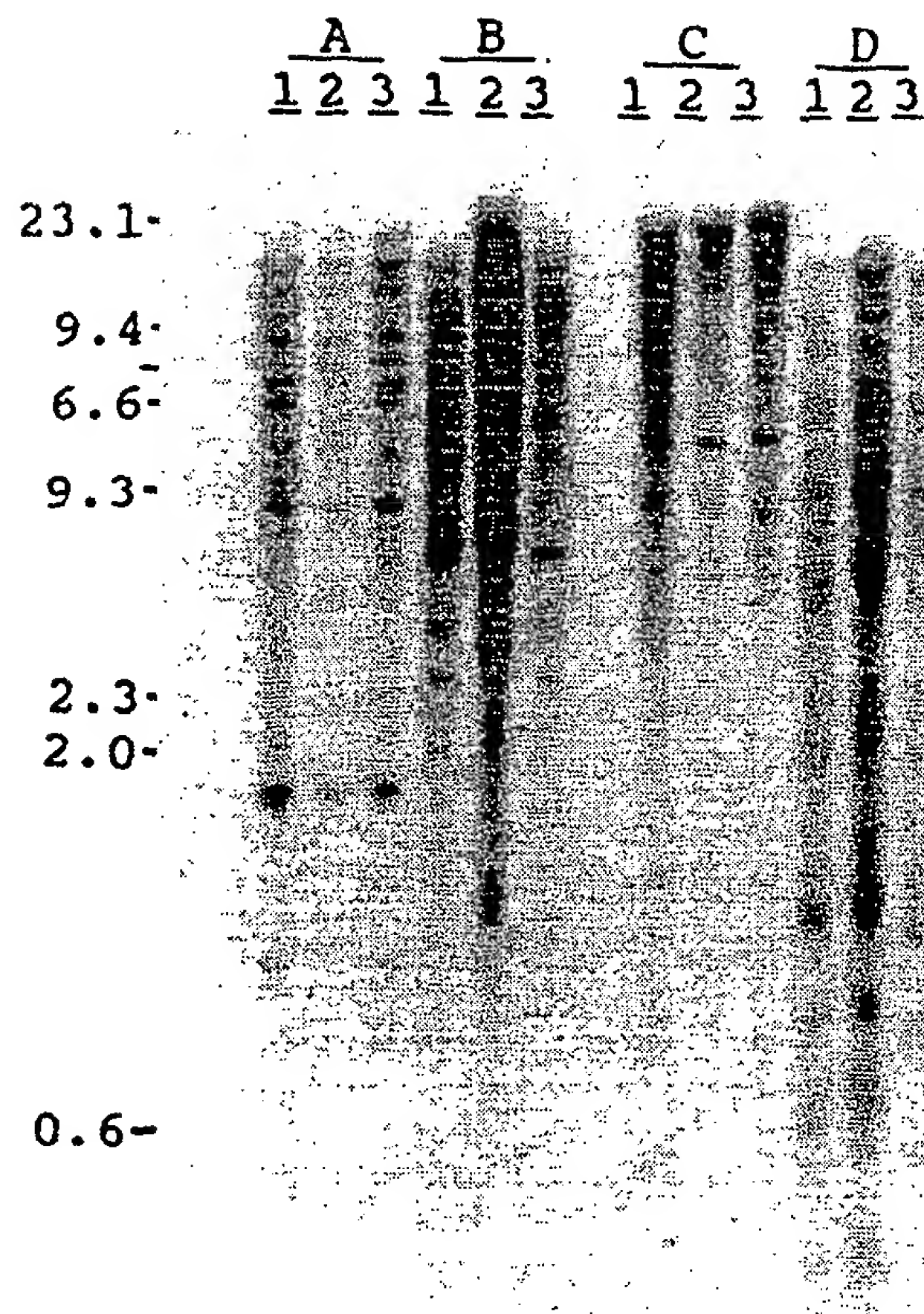
361 GCTGTGCTGACCATCTACTATCGAACAGGCTGATCAGGTACTTCAGGCCCAAGTCCCTT
-----+-----+-----+-----+-----+420
 A V V T I Y Y G T G L I R Y L R P K S L

421 TATTCCGCTGAGGAGACAGACTGATCTCTGTGTTCTATGCAGTCATTGGCCCTGCACTG
-----+-----+-----+-----+-----+480
 Y S A E G D R L I S V F Y A V I G P A L

98/99

SUBSTITUTE SHEET

99/99

Figure 32

SUBSTITUTE SHEET

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/02741

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : C12N 15/12, 15/63, 15/64, 5/10; C07K 13/00; A01N 33/00; A61K 37/00

US CL : 536/27; 424/418; 435/7.21, 172.3, 240.1, 320.1; 514/2; 530/395

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/27; 424/418; 435/7.21, 172.3 240.1, 320.1; 514/2; 530/395

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CAS ONLINE, MEDLINE, UEMBL, GENBANK, PIR, SWISS PROT, APS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P Y,P	Molecular Brain Research, Volume 13, No. 1-2, issued March 1992, L. A. Selbie et al., "Novel G protein-coupled receptors: a gene family of putative human olfactory receptor sequences," abstract.	1-32 33-98
Y X	Sensory Syst., Volume 1, No. 1, issued 1987, V. I. Novoselov et al., "The properties of receptor molecules from rat olfactory epithelium," abstract.	1-34, 65-98 35-64
X,P Y,P	Nature, Volume 355, issued 30 January 1992, M. Parmentier et al., "Expression of members of the putative olfactory receptor gene family in mammalian germ cells," pages 453-455, see entire document.	1-32 33-98
Y X	Biochimica Biophysica Acta, Volume 839, No. 3, issued 1985, E. E. Fesenko et al., "Molecular mechanisms of olfactory reception. VI Kinetic characteristics of camphor interaction with binding sites of rat olfactory epithelium," abstract.	1-34, 65-98 35-64

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

- * Special categories of cited documents:
- *A* document defining the general state of the art which is not considered to be part of particular relevance
- *E* earlier document published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

Date of the actual completion of the international search

25 June 1992

Date of mailing of the international search report

23 July 1992

Name and mailing address of the ISA/
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. NOT APPLICABLE

Authorized officer

LISA T. BENNETT

Telephone No. (703) 308-3988

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/02741

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P Y,P	Chemtracts: Organic Chemistry, Volume 4, No. 4, issued 1991, K. Touhara et al., "A novel multigene family may encode odorant receptors: a molecular basis for odor recognition," abstract.	1-32 33-98
Y,P	Chemical Senses, Volume 16, No. 5, issued 1991, R. H. R. Anholt, "Odor recognition and olfactory transduction: the new frontier," abstract.	1-98
Y	Trends in Neuroscience, Volume 14, No. 7, issued 1991, S. Firestein, "A noseful of odor receptors," abstract.	1-98
Y	Proceedings of the National Academy of Sciences, Volume 86, issued November 1989, E. Dancigier et al., "Olfactory marker protein gene: Its structure and olfactory neuron-specific expression in transgenic mice," pages 8565-8569, see entire document.	1-34
Y	Kagaky Kogyo, Volume 40, No. 11, issued 1989, M. Kashiwayanagi et al., "High sensitivity odor sensor using artificial membrane," abstract.	1-98

Form PCT/ISA/210 (continuation of second sheet)(July 1992)*



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : C12N 15/12, 15/63, 15/64 C12N 5/10, C07K 13/00 A01N 33/00, A61K 37/00	A1	(11) International Publication Number: WO 92/17585 (43) International Publication Date: 15 October 1992 (15.10.92)
(21) International Application Number: PCT/US92/02741 (22) International Filing Date: 6 April 1992 (06.04.92) (30) Priority data: 681,880 5 April 1991 (05.04.91) US (60) Parent Application or Grant (63) Related by Continuation US 681,880 (CIP) Filed on 5 April 1991 (05.04.91) (71) Applicant (for all designated States except US): THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK [US/US]; Broadway and West 116th Street, New York, NY 10027 (US).		(72) Inventors; and (75) Inventors/Applicants (for US only): BUCK, Linda, B. [US/US]; 100 Haven Avenue, New York, NY 10032 (US). AXEL, Richard [US/US]; 445 Riverside Drive, New York, NY 10027 (US). (74) Agent: WHITE, John, P.; Cooper & Dunham, 30 Rockefeller Plaza, New York, NY 10112 (US). (81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), MC (European patent), NL (European patent), SE (European patent), US. Published <i>With international search report.</i>
(54) Title: ODORANT RECEPTORS AND USES THEREOF (57) Abstract The invention provides an isolated nucleic acid, e.g. cDNA encoding an odorant receptor. The invention further provides expression vectors containing such nucleic acid. Also provided is a purified protein encoding an odorant receptor, with the aforementioned expression vectors and the resulting transformed cell. The invention also provides methods of identifying odorant ligands and of identifying odorant receptors. The invention further provides methods of developing fragrances, of identifying appetite suppressant compounds, of controlling appetite, of controlling pest populations, of promoting and inhibiting fertility, and of detecting odors.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	RU	Russian Federation
CG	Congo	KP	Democratic People's Republic of Korea	SD	Sudan
CH	Switzerland	KR	Republic of Korea	SE	Sweden
CI	Côte d'Ivoire	LI	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
DE	Germany	MC	Monaco	TC	Togo
DK	Denmark			US	United States of America

ODORANT RECEPTORS AND USES THEREOF**Background of the Invention**

5 This application is a continuation-in-part of U.S. Serial No.681,880, filed April 5, 1991, the contents of which are hereby incorporated by reference.

10 Throughout this application, various publications are referenced by Arabic numerals within parentheses. Full citations for these publications may be found at the end of the specification immediately preceding the claims. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in
15 order to more fully describe the state of the art as known to those skilled therein as of the date of the invention described and claimed herein.

20 In vertebrate sensory systems, peripheral neurons respond to environmental stimuli and transmit these signals to higher sensory centers in the brain where they are processed to allow the discrimination of complex sensory information. The delineation of the peripheral mechanisms by which environmental stimuli are transduced into neural information
25 can provide insight into the logic underlying sensory processing. Our understanding of color vision, for example, emerged only after the observation that the discrimination of hue results from the blending of information from only three classes of photoreceptors (1, 2, 3, 4). The basic
30 logic underlying olfactory sensory perception, however, has remained elusive. Mammals possess an olfactory system of enormous discriminatory power (5, 6). Humans, for example, are thought to be capable of distinguishing among thousands of distinct odors. The specificity of odor recognition is
35 emphasized by the observation that subtle alterations in the molecular structure of an odorant can lead to profound

-2-

changes in perceived odor.

The detection of chemically distinct odorant presumably results from the association of odorous ligands with specific receptors on olfactory neurons which reside in a specialized epithelium in the nose. Since these receptors have not been identified, it has been difficult to determine how odor discrimination might be achieved. It is possible that olfaction, by analogy with color vision, involves only a few odor receptors, each capable of interaction with multiple odorant molecules. Alternatively, the sense of smell may involve a large number of distinct receptors each capable of associating with one or a small number of odorant. In either case, the brain must distinguish which receptors or which neurons have been activated to allow the discrimination between different odorant stimuli. Insight into the mechanisms underlying olfactory perception is likely to depend upon the isolation of the odorant receptors, and the characterization of their diversity, specificity, and patterns of expression.

The primary events in odor detection occur in a specialized olfactory neuroepithelium located in the posterior recesses of the nasal cavity. Three cell types dominate this epithelium (Figure 1A): the olfactory sensory neuron, the sustentacular or supporting cell, and the basal cell which is a stem cell that generates olfactory neurons throughout life (7, 8). The olfactory sensory neuron is bipolar: a dendritic process extends to the mucosal surface where it gives rise to a number of specialized cilia which provide an extensive, receptive surface for the interaction of odors with olfactory sensory neurons. The olfactory neuron also gives rise to an axon which projects to the olfactory bulb of the brain, the first relay in the olfactory system. The axons of the olfactory bulb neurons, in turn, project to

-3-

subcortical and cortical regions where higher level processing of olfactory information allows the discrimination of odors by the brain.

5 The initial events in odor discrimination are thought to involve the association of odors with specific receptors on the cilia of olfactory neurons. Selective removal of the cilia results in the loss of olfactory response (9). Moreover, in fish, whose olfactory system senses amino acids
10 as odors, the specific binding of amino acids to isolated cilia has been demonstrated (10, 11). The cilia are also the site of olfactory signal transduction. Exposure of isolated cilia from rat olfactory epithelium to numerous odorant leads to the rapid stimulation of adenylyl cyclase
15 and elevations in cyclic AMP (an elevation in IP3 in response to one odorant has also been observed) (12, 13, 14, 15). The activation of adenylyl cyclase is dependent on the presence of GTP and is therefore likely to be mediated by receptor-coupled GTP binding proteins (G-proteins) (16).
20 Elevations in cyclic AMP, in turn, are thought to elicit depolarization of olfactory neurons by direct activation of a cyclic nucleotide-gated, cation permeable channel (17, 18). This channel is opened upon binding of cyclic nucleotides to its cytoplasmic domain, and can therefore
25 transduce changes in intracellular levels of cyclic AMP into alterations in the membrane potential.

These observations suggest a pathway for olfactory signal transduction (Figure 1B) in which the binding of odors to
30 specific surface receptors activates specific G-proteins. The G-proteins then initiate a cascade of intracellular signalling events leading to the generation of an action potential which is propagated along the olfactory sensory axon to the brain. A number of neurotransmitter and hormone
35 receptors which transduce intracellular signals by

-4-

activation of specific G-proteins have been identified. Gene cloning has demonstrated that each of these receptors is a member of a large superfamily of surface receptors which traverse the membrane seven times (19, 20). The
5 pathway of olfactory signal transduction (Figure 1B) predicts that the odorant receptors might also be members of this superfamily of receptor proteins. The detection of odors in the periphery is therefore likely to involve signalling mechanisms shared by other hormone or
10 neurotransmitter systems, but the vast discriminatory power of the olfactory system will require higher order neural processing to permit the perception of individual odors. This invention address the problem of olfactory perception at a molecular level. Eighteen different members of an
15 extremely large multigene family have been cloned and characterized which encodes seven transmembrane domain proteins whose expression is restricted to the olfactory epithelium. The members of this novel gene family encode the individual odorant receptors.

20

-5-

SUMMARY OF THE INVENTION

The invention provides an isolated nucleic acid, e.g. a DNA and cDNA molecule, encoding an odorant receptor. The invention further provides expression vectors containing such nucleic acid. Also provided by the invention is a purified protein encoding an odorant receptor. The invention further provides a method of transforming cells which comprises transfecting a suitable host cell with a suitable expression vector containing the nucleic acid encoding the odorant receptor.

The invention also provides methods of identifying odorant ligands and of identifying odorant receptors. The invention further provides methods of developing fragrances, of identifying appetite suppressant compounds, of controlling appetite. The invention also provides methods of controlling insect and other animal populations. The invention additionally provides a method of detecting odors such as the vapors emanating from Cocaine, Marijuana, Heroin, Hashish, Angel Dust, gasoline, decayed human flesh, alcohol, gun powder explosives, plastic explosives, firearms, poisonous or harmful smoke, or natural gas.

-6-

Description of the Figures

Figure 1. The Olfactory Neuroepithelium and a Pathway for Olfactory Signal Transduction. A. The Olfactory Neuroepithelium. The initial event in odor perception occurs in the nasal cavity in a specialized neuroepithelium which is diagrammed here. Odors are believed to interact with specific receptors on the cilia of olfactory sensory neurons. The signal generated by these initial binding events are propagated by olfactory neuron axons to the olfactory bulb. B. A Pathway of Olfactory Signal Transduction. In this scheme, the binding of an odorant molecule to an odor-specific transmembrane receptor leads to the interaction of the receptor with a GTP-binding protein ($G_{S(olf)}$). This interaction, in turn, leads to the release of the GTP-coupled α -subunit of the G-protein, which then stimulates adenylyl cyclase to produce elevated levels of cAMP. The increase in cAMP opens nucleotide-gated cation channels, thus causing an alteration in membrane potential.

Figure 2. A PCR Amplification Product Containing Multiple Species of DNA. cDNA prepared from olfactory epithelium RNA was subjected to PCR amplification with a series of different primer oligonucleotides and the DNA products of appropriate size were isolated, further amplified by PCR, and size fractionated on agarose gels (A) (For details, see text). Each of these semipurified PCR products was digested with the restriction enzyme, Hinf I, and analyzed by agarose gel electrophoresis. Lanes marked "M" contain size markers of 23.1, 9.4, 5.6, 4.4, 2.3, 2.0, 1.35, 1.08, 0.87, 0.60, 0.31, 0.28, 0.23, 0.19, 0.12 and 0.07kb. (B). Twenty-two of the 64 PCR products that were isolated and digested with Hinf I are shown here. Digestion of one of these, PCR 13, yielded a large number of fragments whose sizes summed to a value much greater than that of the undigested PCR 13

-7-

DNA, indicating that PCR 13 might contain multiple species of DNA which are representatives of a multigene family.

Figure 3. Northern Blot Analysis with a Mixture of Twenty Probes. One μ g of polyA⁺ RNA isolated from rat olfactory epithelium, brain, or spleen was size-fractionated in formaldehyde agarose, blotted onto a nylon membrane, and hybridized with a ³²P-labeled mixture of segments of 20 cDNA clones. The DNA segments were obtained by PCR using primers homologous to transmembrane domains 2 and 7.

Figure 4. The Protein Sequences Encoded by Ten Divergent cDNA Clones. Ten divergent cDNA clones were subjected to DNA sequence analyses and the protein sequence encoded by each was determined. Amino acid residues which are conserved in 60% or more of the proteins are shaded. The presence of seven hydrophobic domains (I-VII), as well as short conserved motifs shared with other members of the superfamily, demonstrate that these proteins belong to the seven transmembrane domain protein superfamily. Motifs conserved among members of the superfamily and the family of olfactory proteins include the GN in TM1 (transmembrane domain 1), the central W of TM4, the Y near the C-terminal end of TM5, and the NP in TM7. In addition, the DRY motif C-terminal to TM3 is common to many members of the G-protein-coupled superfamily. However, all of the proteins shown here share sequence motifs not found in other members of this superfamily and are clearly members of a novel family of proteins.

Figure 5. Positions of Greatest Variability in the Olfactory Protein Family. In this diagram, the protein encoded by cDNA clone I15 is shown traversing the plasma membrane seven times with its N-terminus located extracellularly, and its C-terminus intracellularly. The

-8-

vertical cylinders delineate the seven putative α -helices spanning the membrane. Positions at which 60% or more of the 10 clones shown in Figure 4 share the same residue as I15 are shown as white balls. More variable residues are shown as black balls. The high degree of variability encountered in transmembrane domains III, IV, and V is evident in this schematic.

Figure 6. The Presence of Subfamilies in a Divergent Multigene Family. Partial nucleotide sequences and deduced protein sequences were obtained for 18 different cDNA clones. Transmembrane domain V along with the flanking loop sequences, including the entire cytoplasmic loop between transmembrane domains V and VI, are shown here for each protein. Amino acid residues found in 60% or more of the clones in a given position are shaded (A). This region of the olfactory proteins (particularly transmembrane domain V) appears to be highly variable (see Figure 4). These proteins, however, can be grouped into subfamilies (B,C,D) in which the individual subfamily members share considerable homology in this divergent region of the protein.

Figure 7. Southern Blot Analyses with Non-crosshybridizing Fragments of Divergent cDNAs. Five μ g of rat liver DNA was digested with Eco RI (A) or Hind III (B), electrophoresed in 0.75% agarose, blotted onto a nylon membrane, and hybridized to the 32 P-labeled probes indicated. The probes used were PCR-generated fragments of: 1, clone F9 (identical to F12 in Figure 4); 2, F5; 3, F6; 4, I3; 5, I7; 6, I14; or 7, I15. The lane labeled "1-7" was hybridized to a mixture of the seven probes. The probes used showed either no crosshybridization or only trace crosshybridization with one another. The size markers on the left correspond to the four blots on the left (1-4) whereas the marker positions noted on the right correspond to the four blots on the right

-9-

(5-7, "1-7").

- Figure 8. Northern Blot Analysis with a Mix of Seven Divergent Clones. One μg of polyA⁺ RNA from each of the tissues shown was size-fractionated, blotted onto a nylon membrane, and hybridized with a ³²P-labeled mixture of segments of seven divergent cDNA clones (see Legend to Figure 7).
- Figure 9. The amino acid and nucleic acid sequence of clone F3.
- Figure 10. The amino acid and nucleic acid sequence of clone F5.
- Figure 11. The amino acid and nucleic acid sequence of clone F6.
- Figure 12. The amino acid and nucleic acid sequence of clone F12.
- Figure 13. The amino acid and nucleic acid sequence of clone I3.
- Figure 14. The amino acid and nucleic acid sequence of clone I7.
- Figure 15. The amino acid and nucleic acid sequence of clone I8.
- Figure 16. The amino acid and nucleic acid sequence of clone I9.
- Figure 17. The amino acid and nucleic acid sequence of clone I14.

-10-

Figure 18. The amino acid and nucleic acid sequence of clone I15.

5 Figure 19. The amino acid and nucleic acid sequence of human clone H5.

10 Figure 20. The amino acid and nucleic acid sequence of clone J1, where the reading frame starts at nucleotide position 2.

Figure 21. The amino acid and nucleic acid sequence of clone J2.

15 Figure 22. The amino acid and nucleic acid sequence of clone J4, where the reading frame starts at nucleotide position 2.

20 Figure 23. The amino acid and nucleic acid sequence of clone J7, where the reading frame starts at nucleotide position 2.

25 Figure 24. The amino acid and nucleic acid sequence of clone J8, where the reading frame starts at nucleotide position 2.

Figure 25. The amino acid and nucleic acid sequence of clone J11.

30 Figure 26. The amino acid and nucleic acid sequence of clone J14, where the reading frame starts at nucleotide position 2.

35 Figure 27. The amino acid and nucleic acid sequence of clone J15, where the reading frame starts at nucleotide position 2.

-11-

Figure 28. The amino acid and nucleic acid sequence of clone J16, where the reading frame starts at nucleotide position 2.

5 Figure 29. The amino acid and nucleic acid sequence of clone J17, where the reading frame starts at nucleotide position 2.

10 Figure 30. The amino acid and nucleic acid sequence of clone J19, where the reading frame starts at nucleotide position 2.

15 Figure 31. The amino acid and nucleic acid sequence of clone J20, where the reading frame starts at nucleotide position 2.

20 Figure 32. SOUTHERN BLOT: Five micrograms of DNA isolated from 1. Human placenta, 2. NCI-H-1011 neuroblastoma cells, or 3. CHP 134 neuroblastoma cells were treated with the restriction enzyme A. Eco RI, B. Hind III, C. Bam HI, or D. Pst I, and then electrophoresed on an agarose gel and blotted onto a nylon membrane. The blotted DNA was hybridized to the ³²P-labeled H3/H5 sequence. An autoradiograph of the hybridized blot is shown with the
25 sizes of co-electrophoresed size markers noted in kilobases.

-12-

Detailed Description of the Invention

The invention provides an isolated nucleic acid, e.g. a DNA or cDNA molecule, encoding an odorant receptor. Such a
5 receptor is a receptor which binds an odorant ligand and include but not limited to pheromone receptors. An odorant ligand may include, but is not limited to, molecules which interact with the olfactory sensory neuron, molecules which interact with the olfactory cilia, pheromones, and molecules
10 which interact with structures within the vomeronasal organ.

The invention specifically provides the isolated cDNAs encoding odorant receptors the sequences of which are shown in Figures 9-31. The nucleic acid is most typically a cDNA
15 and encodes an insect, a vertebrate, a fish or a mammalian odorant receptor. The mammalian odorant receptor is preferably a human, rat, mouse or dog receptor. In an embodiment, human odorant receptor cDNA sequence and the correspondent protein is isolated (Figure 19).

20 In another embodiment, pheromone receptors are isolated and shown as clones J1, J2, J4, J7, J8, J11, J14, J15, J16, J17, J19 and J20 (Figures 20-31).

25 The invention further provides expression vectors containing cDNA which encodes odorant receptors. Such expression vectors are well known in the art and include in addition to the nucleic acid the elements necessary for replication and expression in a suitable hosts. Suitable hosts are well
30 known in the art and include without limitation bacterial hosts such as E. coli, animal hosts such as CHO cells, insect cells, yeast cells and like.

The invention also provides purified proteins encoding
35 odorant receptors. Such proteins may be prepared by

-13-

expression of the forementioned expression vectors in suitable host cells and recovery and purification of the receptors using methods well known in the art. Examples of such proteins include those having the amino acid sequences shown in figures 9-31.

The purified protein typically encodes an insect, vertebrate, fish or mammalian odorant receptor. The mammalian odorant receptor may be a human, rat, mouse or dog.

In one embodiment the invention provides a novel purified protein which belong to a class of proteins which have 7 transmembrane regions and a third cytoplasmic loop from the N-terminus which is approximately 17 amino acid long and to nucleic acid molecules encoding such proteins.

The invention provides methods of transforming cells which comprises transfecting a suitable host cell with a suitable expression vector containing nucleic acid encoding of the odorant receptor. Techniques for carrying out such transformations on cells are well known to those skilled in the art. (41,42) Additionally, the resulting transformed cells are also provided by the invention. These transformed cells may be either olfactory cells or non-olfactory cells. One advantage of using transformed non-olfactory cells is that the desired odorant receptor will be the only odorant receptor expressed on the cell's surface.

In order to obtain cell lines that express a single receptor type, standard procedures may be used to clone individual cDNAs or genes into expression vectors and then transfect the cloned sequences into mammalian cell lines. This approach has been used with sequences encoding some other members of the seven transmembrane domain superfamily

-14-

including the 5HT1c serotonin receptor. (43) The cited work illustrates how members of this superfamily transferred into cell lines may generate immortal cell lines that express high levels of the transfected receptor on the cell surface where it will bind ligand and that such abnormally expressed receptor molecules can transduce signals upon binding to ligand.

The invention also provides a method of identifying a desired odorant ligand which comprises contacting transformed non-olfactory cells expressing a known odorant receptor with a series of odorant ligands to determining which ligands bind to the receptors present on the non-olfactory cells.

Additionally, the invention provides a method of identifying a desired odorant receptor comprising contacting a series of transformed non-olfactory cells with a known odorant ligand and determining which odorant receptor binds with the odorant ligand.

The invention provides a method of detecting an odor which comprises: a) identifying a odorant receptor which binds the desired odorant ligand and; b) imbedding the receptor in a membrane such that when the odorant ligand binds to the receptor so identified a detectable signal is produced. In one embodiment of the invention the membrane used in this method is cellular, including a membrane of an olfactory cell or a synthetic membrane.

The ligand tested for may be the vapors emanating from Cocaine, Marijuana, Heroin, Hashish, Angel Dust, gasoline, decayed human flesh, alcohol, gun powder explosives, plastic explosives or firearms. In another embodiment the ligand tested for may be natural gas, a pheromone, toxic fumes,

-15-

noxious fumes or dangerous fumes.

In one embodiment of the invention the detectable signal is a lightbulb lighting up, a buzzer buzzing, a bell ringing,
5 a color change, phosphorescence, or radioactivity.

The invention further provides a method of quantifying the amount of an odorant ligand present in a sample which comprises utilizing the above-mentioned method for odor
10 detection and then quantifying the amount of signal produced.

The invention further provides a method of developing fragrances which comprises identifying a desired odorant
15 receptor by the above method, then contacting non-olfactory cells, which have been transfected with an expression vector containing nucleic acid encoding the desired odorant receptor such that the receptor is expressed upon the surface of the non-olfactory cell, with a series of
20 compounds to determine which compound or compounds bind the receptor.

The invention provides to a method of identifying an "odorant fingerprint" which comprises contacting a series of
25 cells, which have been transformed such that each express a known odorant receptor, with a desired sample and determining the type and quantity of the odorant ligands present in the sample.

30 The invention provides a method of identifying odorant ligands which inhibit the activity of a desired odorant receptor which comprises contacting the desired odorant receptor with a series of compounds and determining which compounds inhibit the odorant ligand - odorant receptor
35 interaction.

-16-

The invention also provides for a method of identifying appetite suppressant compounds which comprises identifying odorant ligands by the method mentioned in the preceding paragraph wherein the desired odorant receptor is that which is associated with the perception of food. Additionally, the invention provides a method of controlling appetite in a subject which comprises contacting the olfactory epithelium of the subject with these odorant ligands. Further the invention provides a nasal spray, to control appetite comprising the compounds identified by the above method in a suitable carrier.

The invention provides a method of trapping odors which comprises contacting a membrane which contains multiples of the desired odorant receptor, with a sample such that the desired odorant ligand is absorbed by the binding of the odorant ligand to the odorant receptor. The invention also provides an odor trap employing this method.

The invention also provides a method of controlling pest populations which comprises identifying odorant ligands by the method mentioned above which are alarm odorant ligands and spraying the desired area with the identified odorant ligands. Additionally, provided by the invention is a method of controlling a pest population which comprises identifying odorant ligands by the above mentioned method, which interfere with the interaction between the odorant ligands and the odorant receptors which are associated with fertility. In one embodiment the pest population is a population of insects or rodents, including mice and rats.

The invention also provides a method of promoting fertility which comprises identifying odorant ligands which interact with the odorant receptors associated with fertility by the above mentioned method. Further, the invention provides a

-17-

method of inhibiting fertility which comprises employing the above mentioned method to identifying odorant ligands which inhibit the interaction between the odorant ligands and the odorant receptors associated with fertility.

5

This invention is illustrated in the Experimental Detail section which follow. These sections are set forth to aid in an understanding of the invention but are not intended to, and should not be construed to, limit in any way the invention as set forth in the claims which follow thereafter.

10

EXPERIMENTAL DETAILS

15

MATERIALS AND METHODS

Polymerase Chain Reaction

RNA was prepared from the olfactory epithelia of Sprague Dawley rats according to Chirgwin et al. (40) or using RNazol B (Cinna/Biotechx) and then treated with DNase I (0.1 unit/ μ g RNA) (Promega). In order to obtain cDNA, this RNA was incubated at 0.1 μ g/ μ l with 5 μ M random hexamers (Pharmacia) 1 mM each of dATP, dCTP, dGTP, TTP, and 2 units/ μ l RNase inhibitor (Promega) in 10 mM TrisCl (pH 8.3), 50 mM KCl, 2.5 mM MgCl₂, and 0.001% gelatin for 10 min. at 22°C, and then for a further 45 min. at 37°C following the addition of 20 u./ μ l of Moloney murine leukemia virus reverse transcriptase (BRL). After heating at 95°C for 3 min., cDNA prepared from 0.2 μ g of RNA was used in each of a series of polymerase chain reactions (PCR) containing 10 mM TrisCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.001% gelatin, 200 μ M each of dATP, dCTP, dGTP, and TTP, 2.5 u. Taq polymerase (Perkin Elmer Cetus), and 2 μ M of each PCR

35

-18-

primer. PCR reactions were performed according to the following schedule: 96°C for 45 sec., 55°C for 4 min. (or 45°C for 2 min.), 72°C for 3 min. with 6 sec. extension per cycle for 48 cycles. The primers used for PCR were a series of degenerate oligonucleotides made according to the amino acid sequences found in transmembrane domain 2 and 7 of a variety of different members of the 7 transmembrane domain protein superfamily (19). The regions used correspond to amino acids number 60-70 and 286-295 of clone I15 (Figure 4). Each of five different 5' primers were used in PCR reactions with each of six different 3' primers. The 5' primers had the sequences:

15 C AC A C CT
A1, AATTGGATICTIGTIAATCTIGCIGTIGCIGCIGA;

C C CA A C C
A2, AATTATTTTCTIGTIAATCTIGCITTIGCIGA;

20 CCA CC A C
A3, AATTTITTTATATITCICTIGCITGIGCIGA;

A T C T ACT C
A4, CGITTICTIATGTGTAACTITGCTTTGCIGA;

25 C CT TG
A5, ACIGTITATATIACICATCTIACIATIGCIGA.

The 3' primers were:

30 TTA T CAG C C A
B1, CTGICGGTTCATIAAIAACATAIATATIGGGTT;

TG GA G G A A
35 B2, GATCGTTIAGACAACAATAIATATIGGGTT;

A G G A
B3, TCIATGTTAAAGTIGTATAIATATIGGGTT;

40 T G G A A
B4, GCCTTIGTAAAIATIGCATAIAGGAAIGGGTT;

G AGA G G G A
B5, AAATCIGGGCTICGICAATAIATCAIIGGGTT;

45 CT CT G G G G A

-19-

B6, GAIGAICCIACAAAAAATAIATAAAIGGGTT.

5 An aliquot of each PCR reaction was analyzed by agarose gel electrophoresis and bands of interest were amplified further by performing PCR reactions on pipet tip (approx. 1 μ l) plugs of the agarose gels containing those DNAs. Aliquots of these semi-purified PCR products were digested with the restriction enzymes Hae III or Hinf I and the digestion
10 products were compared with the undigested DNAs on agarose gels.

Isolation and Analysis of cDNA Clones

15 CDNA libraries were prepared according to standard procedures (41, 42) in the cloning vector, λ ZAP II (Stratagene) using poly A⁺ RNA prepared from Sprague Dawley rat epithelia (see above) or from an enriched population of olfactory neurons which had been obtained by a 'panning'
20 procedure, using an antibody against the H blood group antigen (Chembiomed) found on a large percentage of rat olfactory neurons. In initial library screens, 8.5×10^5 independent clones from the olfactory neuron library and 1.8×10^6 clones from the olfactory epithelium library were
25 screened (41) with a ³²P-labeled probe (prime-it, Stratagene) consisting of a pool of gel-isolated PCR products obtained using primers A4 and B6 (see above) in PCR reactions using as template, olfactory epithelium cDNA, rat liver DNA, or DNA prepared from the two cDNA libraries. In
30 later library screens, a mixture of PCR products obtained from 20 cDNA clones with the A4 and B6 primers was used as probe ('P1' probe). In initial screens, phage clones were analyzed by PCR using primers A4 and B6 and those which showed the appropriate size species were purified. In later
35 screens, all position clones were purified, but only those that could be amplified with the B6 primer and a primer

-20-

specific for vector sequence were analyzed further. To obtain plasmids from the isolated phage clones, phagemid rescue was performed according to the instructions of the manufacturer of λ ZAP II (Stratagene). DNA sequence analysis was performed on plasmid DNAs using the Sequenase system (USB), initially with the A4 and B6 primers and later with oligonucleotide primers made according to sequences already obtained.

10 Northern and Southern Blot Analyses

For Northern blots, poly A⁺ RNAs from various tissues were prepared as described above or purchased from Clontech. One μ g of each RNA was size fractionated on formaldehyde agarose gels and blotted onto nylon membranes (41, 42). For Southern blots, genomic DNA prepared from Sprague Dawley rat liver was digested with the restriction enzymes Eco RI or Hind III, size fractionated on agarose gels and blotted onto nylon membranes (41, 42). The membranes were dried at 80°C, and then prehybridized in 0.5 M sodium phosphate buffer (pH 7.3) containing 1% bovine serum albumin and 4% sodium dodecyl sulfate. Hybridization was carried out in the same buffer at 65°-70°C for 14-20 hrs. with DNAs labeled with ³²P. For the first Northern blot shown, the 'P1' probe (see above under cDNA clone isolation) was used. For the second Northern blot shown, a mix of PCR fragments from seven divergent cDNA clones was used. For Southern blots, the region indicated in clone I15 by amino acids 118 through 251 was amplified from a series of divergent cDNA clones using PCR. The primers used for these reactions had the sequences:

P1, ATGGCITATGATCGITATGTIGC, and

35 P4, AAIAGIGAIACIATIGAIAGATGIGAICC

-21-

These DNAs (or a DNA encompassing transmembrane domains 2 through 7 for clone F6) were labeled and tested for crosshybridization at 70°C. Those DNAs which did not show appreciable crosshybridization were hybridized individually, or as a pool to Southern blots at 70°C.

Rat Sequences used to obtain similar sequences expressed in Humans

There are genes similar to the rat genes discussed above present in humans, these genes may be readily isolated by screening human gene libraries with the cloned rat sequences or by performing PCR experiments on human genomic DNA with primers homologous to the rat sequences. First, PCR experiments were performed with genomic DNA from rat, human, mouse, and several other species. When primers homologous to transmembrane domains 2 and 6 (the A4/B6 primer set used to isolate the original rat sequences) were used, DNA of the appropriate size was amplified from rat, human and mouse DNAs. When these primary PCR reactions were subsequently diluted and subjected to PCR using primers to internal sequences (P1 and P4 primers), smaller DNA species were amplified whose size was that seen when the same primers were used in PCR reactions with the cloned rat cDNAs. Similarly, when the secondary PCR was performed with one outer primer together with one inner primer (ie. A4/P4 or P1/B6), amplified DNAs were obtained whose sizes were also consistent with the amplification of genes similar in sequence and organization to the cloned rat cDNAs. Second, a mix of segments from 20 of the rat cDNAs ("P1" probe) was used to screen libraries constructed from human genomic DNAs. Hybridization under high or low stringency conditions reveals the presence of a large number of cloned human DNA segments that are homologous to the rat sequences. Finally, RNA from a human olfactory tumor (neuroesthesioma,

-22-

NCI-H-1011) cell line has been examined for sequences homologous to those cloned in the rat. cDNA prepared from this RNA was subjected to PCR with the A4/B6 primer set and a DNA species of the appropriate size was seen. This DNA was subcloned and partially sequenced and clearly encodes a member of the olfactory protein family identified in the rat.

The inserted sequence in human clones H3/H5 was amplified by PCR with the A4/B6 primers, gel purified, and then labeled with ³²P. The labeled DNA was then hybridized to restriction enzyme human placenta. Multiple hybridizing species were observed with each DNA (See Figure 32). This observation is consistent with the presence of a family of odorant receptor genes in the human genome.

The sequence of clone H5 is hereby shown in Figure 19. In addition, the translated protein sequence is shown in Figure 19.

In order to identify odorant receptors in other species, degenerated primer oligonucleotides homologous to conserved regions within the rat odorant receptor family may be used in PCR reactions with genomic DNA or with cDNA prepared from olfactory tissue RNA from those species.

RESULTS

Cloning the Gene Family

A series of degenerate oligonucleotides were designated which could anneal to conserved regions of members of the superfamily of G-protein coupled seven transmembrane domain receptor genes. Five degenerate oligonucleotides (A1-5; see Experimental Procedures) matching sequences within transmembrane domain 2, and six degenerate oligonucleotides

-23-

(B1-6) matching transmembrane domain 7 were used in all combinations in PCR reactions to amplify homologous sequences in cDNA prepared from rat olfactory epithelium RNA. The amplification products of each PCR reaction were then analyzed by agarose gel electrophoresis. Multiple bands were observed with each of the primer combinations. The PCR products within the size range expected for this family of receptors (600 to 1300 bp) were subsequently picked and amplified further with the appropriate primer pair in order to isolate individual PCR bands. Sixty-four PCR bands isolated in this fashion revealed only one or a small number of bands upon agarose gel electrophoresis. Representatives of these isolated PCR products are shown in Figure 2A.

15

The isolated PCR products were digested with the endonuclease, Hae III or Hinf I, which recognize four base restriction sites and cut DNA at frequent intervals. In most instances, digestion of the PCR product with Hinf I generated a set of fragments whose molecular weights sum to the size of the original DNA (Figure 2B). These PCR bands are therefore likely to each contain a single DNA species. In some cases, however, restriction digestion yielded a series of fragments whose molecular weights sum to a value greater than that of the original PCR product. The most dramatic example is shown in Figure 2 where the 710 bp, PCR 13 DNA, is cleaved by Hinf I to yield a very large number of restriction fragments whose sizes sum to a value five- to ten-fold greater than that of the original PCR product. These observations indicated that PCR product 13 consists of a number of different species of DNA, each of which could be amplified with the same pair of primer oligonucleotides. In addition, when PCR experiments similar to those described were performed using cDNA library DNAs as templates, a 710 bp PCR product was obtained with the PCR13 primer pair

-24-

(A4/B6) with DNA from olfactory cDNA libraries, but not a glioma cDNA library. Moreover, digestion of one of this 710 bp product also revealed the presence of multiple DNA species. In other cases (see PCR product 20, for example),
5 digestion yielded a series of restriction fragments whose molecular weights also sum to a size greater than the starting material. Further analysis, however, revealed that the original PCR product consisted of multiple bands of similar but different sizes.

10

In order to determine whether the multiple DNA species present in PCR 13 encode members of a family of seven transmembrane domain proteins, PCR 13 DNA was cloned into the plasmid vector Bluescript and five individual clones
15 were subjected to DNA sequence analysis. Each of the five clones exhibited a different DNA sequence, but each encoded a protein which displayed conserved features of the superfamily of seven transmembrane domain receptor proteins. In addition, the proteins encoded by all five clones shared
20 distinctive sequence motifs not found in other superfamily members indicating they were all members of a new family of receptors.

To obtain full-length cDNA clones, cDNA libraries prepared
25 from olfactory epithelium RNA or from RNA of an enriched population of olfactory sensory neurons were screened. The probe used in these initial screens was a mixture of PCR 13 DNA as well as DNA obtained by amplification of rat genomic DNA or DNA from two olfactory cDNA libraries with the same
30 primers used to generate PCR 13 (A4 and B6 primers). Hybridizing plaques were subjected to PCR amplification with the A4/B6 primer set and only those giving a PCR product of the appropriate size (approximately 710 bp) were purified. The frequency of such positive clones in the enriched
35 olfactory neuron cDNA library was approximately five times

-25-

greater than the frequency in the olfactory epithelium cDNA library. The increased frequency of positive clones observed in the olfactory neuron library is comparable to the enrichment in olfactory neurons generally obtained in the purification procedure.

The original pair of primers used to amplify PCR 13 DNA were then used to amplify coding segments of 20 different cDNA clones. A mix of these PCR products were labeled and used as probe for further cDNA library screens. This mixed probe was also used in a Northern blot (Figure 3) to determine whether the expression of the gene family is restricted to the olfactory epithelium. The mixed probe detects two diffuse bands centered at 2 and 5 kb in RNA from olfactory epithelium; no hybridization can be detected in brain or spleen. (Later experiments which examined a larger number of tissue RNAs with a more restricted probe will be shown below.) Taken together, these data indicate the discovery of a novel multigene family encoding seven transmembrane domain proteins which are expressed in olfactory epithelium, and could be expressed predominantly or exclusively in olfactory neurons.

The Protein Sequences of Numerous, Olfactory-specific Members of the Seven Transmembrane Domain Superfamily

Numerous clones were obtained upon screening cDNA libraries constructed from olfactory epithelium and olfactory neuron RNA at high stringency. Partial DNA sequences were obtained from 36 clones; 18 of these cDNA clones are different, but all of them encode proteins which exhibit shared sequence motifs indicating that they are members of the family identified in PCR 13 DNA. A complete nucleotide sequence was determined for coding regions of ten of the most divergent clones (Figure 4). The deduced protein sequences

-26-

of these cDNAs defines a new multigene family which shares sequence and structural properties with the superfamily of neurotransmitter and hormone receptors that traverse the membrane seven times. This novel family, however, exhibits
5 features different from any other member of the receptor superfamily thus far identified.

Each of the ten sequences contains seven hydrophobic stretches (19-26 amino acids) that represent potential
10 transmembrane domains. These domains constitute the regions of maximal sequence similarity to other members of the seven transmembrane domain superfamily (see legend to Figure 4). On the basis of structural homologies with rhodopsin and the β -adrenergic receptors, (19) it is likely that the amino
15 termini of the olfactory proteins are located on the extracellular side of the plasma membrane and the carboxyl termini are located in the cytoplasm. In this scheme, three extracellular loops alternate with three intracellular loops to link the seven transmembrane domains (see Figure 5).
20 Analysis of the sequences in figure 4 demonstrates that the olfactory proteins, like other members of the receptor superfamily, display no evidence of an N-terminal signal sequence. As in several other superfamily members, a potential N-linked glycosylation site is present in all ten
25 proteins within the short N-terminal extracellular segment. Other structural features conserved with previously identified members of the superfamily included cysteine residues at fixed positions within the first and second extracellular loops that are thought to form a disulfide
30 bond. Finally, many of the olfactory proteins reveal a conserved cysteine within the C-terminal domain which may serve as a palmitoylation site anchoring this domain to the membrane (21). These features, taken together with several short, conserved sequence motifs (see legend to Figure 4),
35 clearly define this new family as a member of the

-27-

superfamily of genes encoding the seven transmembrane domain receptors.

5 There are, however, important differences between the
olfactory protein family and the other seven transmembrane
domain proteins described previously and these differences
may be relevant to proposed function of these proteins in
odor recognition. Structure-function experiments involving
10 in vitro mutagenesis suggest that adrenergic ligands
interact with this class of receptor molecule by binding
within the plane of the membrane (22, 20). Not
surprisingly, small receptor families that bind the same
class of ligands, such as the adrenergic and muscarinic
acetylcholine receptor families exhibit maximum sequence
15 conservation (often over 80%) within the transmembrane
domains. In contrast, the family of receptors discussed in
this application shows striking divergence within the third,
fourth, and fifth transmembrane domains (Figure 4). The
variability in the three central transmembrane domains is
20 highlighted schematically in Figure 5. The divergence in
potential ligand binding domains is consistent with the idea
that the family of molecules cloned is capable of
associating with a large number of odorant of diverse
molecular structure.

25 Receptors which belong to the superfamily of seven
transmembrane domain proteins interact with G-proteins to
generate intracellular signals. In vitro mutagenesis
experiments indicate that one site of association between
30 receptor and G-protein resides within the third cytoplasmic
loop (22, 23). The sequence of this cytoplasmic loop in 18
different clones we have characterized is shown in Figure
6A. This loop which is often quite long and of variable
length in the receptor superfamily is relatively short (only
35 17 amino acids) and of fixed length in the 18 clones

-28-

examined. Eleven of the 18 different clones exhibit the sequence motif K/R I V S S I (or a close relative) at the N-terminus of this loop. Two of the cDNA clones reveal a different H I T C/W A V motif at this site. If this short loop is a site of contact with G-proteins, it is possible that the conserved motifs may reflect sites of interaction with different G-proteins that activate different intracellular signalling systems in response to odors. In addition, the receptors cloned reveal several serine or threonine residues within the third cytoplasmic loop. By analogy with other G-protein coupled receptors, these residues may represent sites of phosphorylation for specific receptor kinases involved in desensitization. (24)

15 Subfamilies within the Multigene Family

Figure 6A displays the sequences of the fifth transmembrane domain and the adjacent cytoplasmic loop encoded by L8 of the cDNA clones we have analyzed. As a group, the 18 sequences exhibit considerable divergence within this region. The multigene family, however, can be divided into subfamilies such that the members of a given subfamily share significant sequence conservation. In subfamily B, clones F12 and F13, for example, differ from one another at only four of 44 positions (91% identity), and clearly define a subfamily. Clones F5 and I11 (subfamily D) differ from F12 and F13 at 34-36 positions within this region and clearly define a separate subfamily. Thus, this olfactory-specific multigene family consists of highly divergent subfamilies. If these genes encode odor receptors, it is possible that members of the divergent subfamilies bind odorant of widely differing structural classes. Members of the individual subfamilies could therefore recognize more subtle differences between molecules which belong to the same structural class of molecules structures.

-29-

The Size of the Multigene Family

Genomic Southern blotting experiments were performed and genomic libraries were screened to obtain an estimate of the sizes of the multigene family and the member subfamilies encoding the putative odor receptors. DNAs extending from the 3' end of transmembrane domain 3 to the middle of transmembrane domain 6 were synthesized by PCR from DNA of seven of the divergent cDNA clones (Figure 4). In initial experiments, these DNAs were labeled and hybridized to each other to define conditions under which minimal crosshybridization would be observed among the individual clones. At 70°C, the seven DNAs showed no crosshybridization, or crosshybridized only very slightly. The trace levels of crosshybridization observed are not likely to be apparent upon genomic Southern blot analysis where the amounts of DNA are far lower than in the test cross.

Probes derived from these seven DNAs were annealed under stringent conditions, either individually or as a group, to Southern blots of rat liver DNA digested with the restriction endonucleases Eco RI or Hind III (Figure 7). Examination of the Southern blots reveals that all but one of the cDNAs detects a relatively large, distinctive array of bands in genomic DNA. Clone I15 (probe 7), for example, detects about 17 bands with each restriction endonuclease, whereas clone F9 (probe 1) detects only about 5-7 bands with each enzyme. A single band is obtained with clone I7 (probe 5). PCR experiments using nested primers (TM2/TM7 primers followed by primers to internal sequences) and genomic DNA as template indicate that the coding regions of the members of this multigene family, like those of many members of the G-protein coupled superfamily, may not be interrupted by introns. This observation, together with the fact that most

-30-

of the probes only encompasses 400 nucleotides suggests that each band observed in these experiments is likely to represent a different gene. These data suggest that the individual probes chosen are representatives of subfamilies which range in size from a single member to as many as 17 members. A total of about 70 individual bands were detected in this analysis which could represent the presence of at least 70 different genes. Although the DNA probes used in these blots did not crosshybridize appreciably with each other, it is possible that a given gene might hybridize to more than one probe, resulting in an overestimate of gene number. However, it is probable that the total number of bands only reflects a minimal estimate of gene number since it is unlikely that we have isolated representative cDNAs from all of the potential subfamilies and the hybridizations were performed under conditions of very high stringency.

A more accurate estimate of the size of the olfactory-specific gene family was obtained by screening rat genomic libraries. The mix of the seven divergent probes used in Southern blots, or the mix of 20 different probes used in our initial Northern blots (see Figure 3), were used as hybridization probes under high (65°C) or lowered (55°C) stringency conditions in these experiments. Nested PCR (see above) was used to verify that the clones giving a positive signal under low stringency annealing conditions were indeed members of this gene family. It is estimated from these studies that there are between 100 and 200 positive clones per haploid genome. The estimate of the size of the family obtain from screens of genomic libraries again represents a lower limit. Given the size of the multigene family, one might anticipate that many of these genes are linked such that a given genomic clone may contain multiple genes. Thus the data from Southern blotting and screens of genomic libraries indicate that the multigene family identified

-31-

consists of one to several hundred member genes which can be divided into multiple subfamilies.

5 It should be noted that the cDNA probes isolated may not be representative of the full complement of subfamilies within the larger family of olfactory proteins. The isolation of cDNAs, for example, relies heavily on PCR with primers from transmembrane domains 2 and 7 and biases our clones for homology within these regions. Thus, estimates of gene
10 number as well as subsequent estimates of RNA abundance should be considered as minimal.

Expression of the Members of this Multigene Family

15 Additional Northern blot analyses were performed to demonstrate that expression of the members of this gene family is restricted to the olfactory epithelium. (Figure 8) Northern blot analysis with a mixed probe consisting of the seven divergent cDNAs used above reveals two diffuse bands
20 about 5 and 2 kb in length in olfactory epithelium RNA. This pattern is the same as that seen previously with the mix of 20 DNAs. No annealing is observed to RNA from the brain or retina or other, nonneural tissues, including lung, liver, spleen, and kidney.

25 An estimate of the level of expression of this family can be obtained from screens of cDNA libraries. The frequency of positive clones in cDNA libraries made from olfactory epithelium RNA suggests that the abundance of the RNAs in
30 the epithelium is about one in 20,000. The frequency of positive clones is approximately five-fold higher in a cDNA library prepared from RNA from purified olfactory neurons (in which 75% of the cells are olfactory neurons). The increased frequency of positive clones obtained in the
35 olfactory neuron cDNA library is comparable to the

-32-

enrichment we obtain upon purification of olfactory neurons. These observations suggest that this multigene family is expressed largely, if not solely, in olfactory neurons and may not be expressed in other cell types within the epithelium. If each olfactory neuron contains 10^5 mRNA molecules, from the frequency of positive clones we predict that each neuron contains only 25-30 transcripts derived from this gene family. Since the family of olfactory proteins consists of a minimum of a hundred genes, a given olfactory neuron could maximally express only a proportion of the many different family members. These values thus suggest that olfactory neurons will exhibit significant diversity at the level of expression of these olfactory proteins.

15

Identification of pheromone receptors in vomeronasal organ

The vomeronasal organ (vomeronasal gland) is an accessory olfactory structure that is located near the nasal cavity. Like the olfactory epithelium of the nasal cavity, the olfactory epithelium of the vomeronasal organ contains olfactory sensory neurons. The vomeronasal organ is believed to play an important role in the sensing of pheromones in numerous species. Pheromones are believed to have profound effects on both physiological and behavioral aspects of reproduction. the identification of pheromone receptors would permit the identification of the pheromones themselves. It would also enable one to identify agonists or antagonists that would either mimic the pheromones or block the pheromone receptors from transducing pheromone signals. Such information would be important to the development of species specific pesticides and, conversely, to animal husbandry. The identification of pheromone receptors in human could ultimately lead to the development of contraceptives or to treatments for infertility in humans. It is likely that the identification of pheromone receptors

-33-

in low mammals such as rodents would lead to the identification of similar receptors in human.

5 In order to identify potential pheromone receptors, we isolate RNA from the vomeronasal organs of female rats and prepared cDNA from this RNA. The cDNA was subjected to PCR with several different pairs of degenerate oligonucleotide primers that match sequences present in the rat odorant receptor family. The PCR products were subcloned and the
10 nucleotide sequences of the subcloned DNAs were determined. Each of the subcloned DNAs encodes a protein that belongs to the odorant receptor family. The sequences of the following vomeronasal subclones are shown: J1, J2, J4, J7, J8, J11, J14, J15, J16, J17, J19, J20. In a few cases (J2, J4), the
15 same sequence was amplified with two different primer pairs and the sequence shown is a composite of the two sequences. It is possible that one or more of these molecules, or closely related molecules, serve as pheromone receptors in the rat.

20

DISCUSSION

The mammalian olfactory system can recognize and discriminate a large number of odorous molecules.
25 Perception in this system, as in other sensory systems, initially involves the recognition of external stimuli by primary sensory neurons. This sensory information is then transmitted to the brain where it is decoded to permit the discrimination of different odors. Elucidation of the logic
30 underlying olfactory perception is likely to require the identification of the specific odorant receptors, the analysis of the extent of receptor diversity and receptor specificity, as well as an understanding of the pattern of receptor expression in the olfactory epithelium.

35

-34-

The odorant receptors are thought to transduce intracellular signals by interacting with G-proteins which activate second messenger systems (12, 13, 14, 15). These proteins are clearly members of the family of G-protein coupled receptors which traverse the membrane seven times (19). The odorant receptors should be expressed specifically in the tissue in which odorant are recognized. The family of olfactory proteins cloned is expressed in the olfactory epithelium. Hybridizing RNA is not detected in brain or retina, or in a host of nonneural tissues. Moreover, expression of this gene family the epithelium may be restricted to olfactory neurons. The family of odorant receptors must be capable of interacting with extremely diverse molecular structures. The genes cloned are members of any extremely large multigene family which exhibit variability in regions thought to be important in ligand binding. The possibility that each member of this large family of seven transmembrane proteins is capable of interacting with only one or a small number of odorant provides a plausible mechanism to accommodate the diversity of odor perception. The properties of the gene family identified suggests that this family is likely to encode a large number of distinct odorant receptors.

25 Size of the Multigene Family

The size of the receptor repertoire is likely to reflect the range of detectable odors and the degree of structural specificity exhibited by the individual receptors. It has been estimated that humans can identify over 10,000 structurally-distinct odorous ligands. However, this does not necessarily imply that humans possess an equally large repertoire of odorant receptors. For example, binding studies in lower vertebrates suggest that structurally-related odorant may activate the same receptor molecules.

-35-

In fish which smell amino acids, the binding of alanine to isolated cilia can be competed by other small polar residues (threonine and serine), but not by the basic amino acids, lysine or arginine (11). These data suggest that individual
5 receptors are capable of associating with several structurally-related ligands, albeit with different affinities. Stereochemical models of olfactory recognition in mammals (25) (based largely on psychophysical, rather than biophysical data) have suggested existence of several
10 primary odor groups including camphoraceous, musky, peppermint, ethereal, pungent, and putrid. In such a model, each group would contain odorant with common molecular configurations which bind to common receptors and share similar odor qualities.

15 Screens of genomic libraries with mixed probes consisting of divergent family members detect approximately 100 to 200 positive clones per genome. The present estimate of at least 100 genes provides only a lower limit since it is
20 likely that the probes used do not detect all of the possible subfamilies. Moreover, it is probable that many of these genes are linked such that a given genomic clone may contain multiple genes. It is therefore expected that the actual size of the gene family may be considerably higher
25 and this family of putative odorant receptors could constitute one of the largest gene families in the genome.

The characterization of a large multigene family encoding putative odorant receptors suggests that the olfactory
30 system utilizes a far greater number of receptors than the visual system. Color vision, for example, allows the discrimination of several hundred hues, but is accomplished by only three different photoreceptors (1, 2, 3 and 4). The photoreceptors each have different, but overlapping
35 absorption spectra which cover the entire spectrum of

-36-

visible wavelengths. Discrimination of color results from comparative processing of the information from these three classes of photoreceptors in the brain. Whereas three photoreceptors can absorb light across the entire visible spectrum, our data suggest that a small number of odorant receptors cannot recognize and discriminate the full spectrum of distinct molecular structures perceived by the mammalian olfactory system. Rather, olfactory perception probably employs an extremely large number of receptors each capable of recognizing a small number of odorous ligands.

Diversity within the Gene Family and the Specificity of Odor Recognition

The olfactory proteins identified in this application are clearly members of the superfamily of receptors which traverse the membrane seven time. Analysis of the proteins encoded by the 18 distinct cDNAs we have cloned reveals structural features which may render this family particularly well suited for the detection of a diverse array of structurally distinct odorant. Experiments with other members of this class of receptors suggest that the ligand binds to its receptor within the plane of the membrane such that the ligand contacts many, if not all of the transmembrane helices. The family of olfactory proteins can be divided into several different subfamilies which exhibit significant sequence divergence within the transmembrane domains. Nonconservative changes are commonly observed within blocks of residues in transmembrane regions 3, 4, and 5 (Figures 4, 5, 6); these blocks could reflect the sites of direct contact with odorous ligands. Some members, for example, have acidic residues in transmembrane domain 3, which in other families are thought to be essential for binding aminergic ligands (20) while other members maintain hydrophobic residues at these positions.

-37-

This divergence within transmembrane domains may reflect the fact that the members of the family of odorant receptors must associate with odorant of widely different molecular structures.

5

These observations suggest a model in which each of the individual subfamilies encode receptors which bind distinct structural classes of odorant. Within a given subfamily, however, the sequence differences are far less dramatic and are often restricted to a small number of residues. Thus, the members of a subfamily may recognize more subtle variations among odor molecules of a given structural class. At a practical level, individual subfamilies may recognize grossly different structures such that one subfamily may associate, for example, with the aromatic compound, benzene and its derivatives, whereas a second subfamily may recognize odorous, short chain, aliphatic molecules. Subtle variations in the structure of the receptors within, for example, the hypothetical benzene subfamily could facilitate the recognition and discrimination of various substituted derivatives such as toluene, xylene or phenol. It should be noted that such a model, unlike previous stereochemical models, does not necessarily predict that molecules with similar structures will have similar odors. The activation of distinct receptors with similar structures could elicit different odors, since perceived odor will depend upon higher order processing of primary sensory information.

30

Evolution of the Gene Family and the Generation of Diversity

35

Preliminary evidence from PCR analyses suggests that members of this family of olfactory proteins are conserved in lower vertebrates as well as invertebrates. This gene family presumably expanded over evolutionary time providing mammals with the ability to recognize an increasing diversity of

-38-

odorant. Examination of the sequences of the family members cloned from mammals provides some insight into the evolution of this multigene family. Although the chromosomal loci encoding these genes has yet to be characterized, it is likely that at least some member genes will be tandemly arranged in a large cluster as is observed with other large multigene families. A tandem array of this sort provides a template for recombination events including unequal crossing over and gene conversion, that can lead to expansion and further diversification of the sort apparent among the family members we have cloned (26).

The multigene family encoding the olfactory proteins is large: all of the member genes clearly have a common ancestral origin, but have undergone considerable divergence such that individual genes encode proteins that share from 40-80% amino acid identity. Subfamilies are apparent with groups of genes sharing greater homology among themselves than with members of other subfamilies. Examination of the sequences of even the most divergent subfamilies, however, reveals a pattern in which several blocks of conserved residues are interspersed with variable regions. This segmental homology is conceptually similar to the organization of framework and hypervariable domains within the families of immunoglobulin and T cell receptor variable region sequences (27, 28). This analogy goes beyond structural organization and may extend to the function of these two gene families: each family consists of a large number of genes which have diversified over evolutionary time to accommodate the binding of a highly diverse array of ligands. The evolutionary mechanisms responsible for the diversification and maintenance of these large gene families may also be similar. It has been suggested that gene conversion has played a major role in the evolution of immunoglobulin and T cell receptor variable domains (29, 30).

-39-

- and 31). Analysis of the sequence of the putative olfactory receptors reveals at least one instance where a motif from a variable region of one subfamily is found imbedded in the otherwise divergent sequence of a second subfamily, suggesting that conversion has occurred. Such a mixing of motifs from one subfamily to another over evolutionary time would provide additional combinatorial possibilities leading to the generation of diversity.
- It should be noted, however, that the combinatorial joining of gene segments by DNA rearrangement during development, which is characteristic of immunoglobulin loci (27), is not a feature of the putative odor receptor gene family. No evidence for DNA rearrangement to generate the diversity of genes cloned has been observed. The entire coding region has been sequenced along with parts of the 5' and 3' untranslated regions of 10 different cDNA clones. The sequences of the coding regions are all different; no evidence has been obtained for constant regions that would suggest DNA rearrangement of the sort seen in the immune system. The observations indicate that the diversity olfactory proteins are coded by a large number of distinct gene sequences.
- Although it is unlikely from the data that DNA rearrangement is responsible for the generation of diversity among the putative odorant receptors, it remains possible that DNA rearrangements may be involved in the regulation of expression of this gene family. If each olfactory neuron expresses only one or a small number of genes, then a transcriptional control mechanism must be operative to choose which of the more than one hundred genes within the family will be expressed in a given neuron. Gene conversion from one of multiple silent loci into a single active locus, as observed for the trypanosome-variable surface

-40-

glycoproteins (32), provides one attractive model. The gene conversion event could be stochastic, such that a given neuron could randomly express any one of several hundred receptor genes, or regulated (perhaps by positional information), such that a given neuron could only express one or a small number of predetermined receptor types. Alternatively, it is possible that positional information in the olfactory epithelium controls the expression of the family of olfactory receptors by more classical mechanisms that do not involve DNA rearrangement. What ever mechanisms will regulate the expression of receptor genes within this large, multigene family, these mechanisms must accommodate the requirement that olfactory neurons are regenerated every 30-60 days (8) and therefore the expression of the entire repertoire of receptors must be accomplished many times during the life of an organism.

Receptor Diversity and the Central Processing of Olfactory Information

The results suggest the existence of a large family of distinct odorant receptors. Individual members of this receptor family are likely to be expressed by only a small set of the total number of olfactory neurons. The primary sensory neurons within the olfactory epithelium will therefore exhibit significant diversity at the level of receptor expression. The question then emerges as to whether neurons expressing the same receptors are localized in the olfactory epithelium. Does the olfactory system employ a topographic map to discriminate among the numerous odorant? The spatial organization of distinct classes of olfactory sensory neurons, as defined by receptor expression, can now be determined by using the procedures of in situ hybridization and immunohistochemistry with probes specific for the individual receptor subtypes. This

-41-

information should help to distinguish between different models that have been proposed to explain the coding of diverse odorant stimuli (33).

5 In one model, sensory neurons that express a given receptor and respond to a given odorant may be localized within defined positions within the olfactory epithelium. This topographic arrangement would also be reflected in the projection of olfactory sensory axons into discrete regions
10 (glomeruli) within the olfactory bulb. In this scheme, the central coding to permit the discrimination of discrete odorant would depend, in part, on the spatial segregation of different receptor populations. Attempts to discern the topographic localization of specific receptors at the level
15 of the olfactory epithelium has led to conflicting results. In some studies, electrophysiological recordings have revealed differences in olfactory responses to distinct odorant in different regions of the olfactory epithelium (34, 35). However, these experiments have been difficult to
20 interpret since the differences in response across the epithelium are often small and are not observed in all studies (36).

A second model argues that sensory neurons expressing
25 distinct odorant receptors are randomly distributed in the epithelium but that neurons responsive to a given odorant project to restricted regions within the olfactory bulb. In this instance, the discrimination of odors would be a consequence of the position of second order neurons in the
30 olfactory bulb, but would be independent of the site of origin of the afferent signals within the epithelium. Mapping of the topographic projections of olfactory neurons has been performed by extracellular recordings from different regions of the bulb (37, 38) and by 2-deoxyglucose
35 autoradiography to map regional activity after exposure to

-42-

different odorant (39). These studies suggest that spatially-localized groups of bulbar neurons preferentially respond to different odorant. The existence of specific odorant receptors, randomly distributed through the olfactory epithelium, which converge on a common target within the olfactory bulb, would raise additional questions about the recognition mechanisms used to guide these distinct axonal subsets to their central targets.

Other sensory systems also spatially segregate afferent input from primary sensory neurons. The spatial segregation of information employed, for example, by the visual and somatosensory systems, is used to define the location of the stimulus within the external environment as well as to indicate the quality of the stimulus. In contrast, olfactory processing does not extract spatial features of the odorant stimulus. Relieved of the necessity to encode information about the spatial localization of the sensory stimulus, it is possible that the olfactory system of mammals uses the spatial segregation of sensory input solely to encode the identity of the stimulus itself. The molecular identification of the genes likely to encode a large family of olfactory receptors should provide initial insights into the underlying logic of olfactory processing in the mammalian nervous system.

-43-

REFERENCES

1. Rushton, W.A.H. (1965). A Foveal pigment in the
5 deuteranope. J. Physiol. 176, 24-37.
2. Rushton, W.A.H. (1955). Foveal photopigments in normal
and colour-blind. J. Physiol. 129, 41-42.
- 10 3. Wald, G., Brown, P.K., and Smith, P.H. (1955).
Iodopsin. J. Gen. Physiol. 38, 623-681.
4. Nathans, J., Thomas, D., and Hogness, D.S. (1986).
15 Molecular genetics of human color vision: The genes
encoding blue, green and red pigments. Science 232,
193-202.
5. Lancet, D. (1986). Vertebrate olfactory reception.
Ann. Rev. Neurosci. 9, 329-355.
20
6. Reed, R.R. (1990). How does the nose know?
Minireview. Cell 60, 1-2.
7. Moulton, D.G., and Beidler, L.M. (1967). Structure and
25 function in the peripheral olfactory system. Physiol.
Rev. 47, 1-52.
8. Graziadei, P.P.C., and Monti Graziadei, G.A. (1979).
Neurogenesis and neuron regeneration in the olfactory
30 system of mammals. I. Morphological aspects of
differentiation and structural organization of the
olfactory sensory neurons. J. Neurocytol. 8, 1-18.
9. Bronshtein, A.A., and Minor, A.V. (1977). Reg neration
35 of olfactory flagella and restoration of the

-44-

electroolfactogram following application of triton X-100 to the olfactory mucosa of frogs. *Tsitologiya* 19, 33-39.

- 5 10. Rhein, L.D., and Cagan, R.H. (1980). Biochemical studies of olfaction: Isolation, characterization, and odorant binding activity of cilia from rainbow trout olfactory rosettes. *Proc. Natl. Acad. Sci. USA* 77, 4412-4416.
- 10 11. Rhein, L.D., and Cagan, R.H. (1983). Biochemical studies of olfaction: Binding specificity of odorant to a cilia preparation from rainbow trout olfactory rosettes. *J. Neurochem.* 41, 569-577.
- 15 12. Pace, U., Hanski, E., Salomon, Y., and Lancet, D. (1985). Odorant-sensitive adenylate cyclase may mediate olfactory reception. *Nature* 316, 255-258.
- 20 13. Sklar, P.B., Anholt, R.R.H., and Snyder, S.H. (1986). The odorant-sensitive adenylate cyclase of olfactory receptor cells: Differential stimulation by distinct classes of odorant. *J. Biol. Chem.* 261, 15538-15543.
- 25 14. Breer, H., Boekhoff, I., and Tareilus, E. (1990). Rapid kinetics of second messenger formation in olfactory transduction. *Nature* 345, 65-68.
- 30 15. Boekhoff, I., Tareilus, E., Strotmann, J., and Breer, H. (1990). Rapid activation of alternative second messenger pathways in olfactory cilia from rats by different odorant. *EMBO J.* 9, 2453
- 35 16. Jones, D.T., and Reed, R.R. (1989). G_{olf} : an olfactory neuron specific-G-protein involved in odorant signal

-45-

transduction. *Science* 244, 790-795.

- 5 17. Nakamura, T. and Gold G. (1987). A cyclic nucleotide-gated conductance in olfactory receptor cilia. *Nature* 325, 442-444.
- 10 18. Dhallan, R.S., Yau, K.-W., Schrader, K.A., and Reed, R.R. (1990). Primary structure and functional expression of a cyclic nucleotide-activated channel from olfactory neurons. *Nature* 347, 184-187.
- 15 19. O'Dowd, B.F., Felkowitz, R.J., and Caron, M.G. (1989). Structure of the adrenergic and related receptors. *Ann. Rev. Neurosci.* 12, 67-83.
- 20 20. Strader, C.D., Sigal, I.S., and Dixon, R.F. (1989). Structural basis of β -adrenergic receptor function *FASEB J.* 3, 1825-1832.
- 25 21. O'Down, B.F., Hnatowich, M., Caron, M.G., Lefkowitz, R.J., and Bouvier, M. (1989) Palmitoylation of the human β_2 -adrenergic receptor. Mutation of CYS³⁴¹ in the carboxyl tail leads to an uncoupled nonpalmytoylated form of the receptor. *J. Biol. Chem.* 264, 7564-7569.
- 30 22. Kobilka, B.KI., Kobilka, T.S., Daniel, K., Regan, J.W., Caron, M.G., and Lefkowitz, R.J. (1988) Chimeric α_2 - β_2 -ldrenergic receptors: delineation of domains involved in effector coupling and ligand binding specificity. *Science* 240, 1310-1316.
- 35 23. Hamm, H.E., Deretic, D., Arendt, A., Hargrove, P.A., Koenig, B., and Hofmann, K.P. (1988) Site of G protein binding to rhodopsin mapping with synthetic peptides to

-46-

the α subunit. *Science* 241, 832-835. (1988).

24. Bouvier, M.W., Hausdorff, A., De Blasi, A., O'Dowd, B.F., Kobilka, B.K., Caron, M.G., and Lefkowitz, R.J. (1988). Removal of phosphorylation sites from the β -adrenergic receptor delays the onset of agonist-promoted desensitization. *Nature (London)* 333, 370-373.
25. Amore, J.E. (1982). Odor Theory and Odor Classification, In *Fragrance Chemistry* (New York: Academic Press, Inc.), pp. 27-73.
26. Maeda, N., and Smithies, O. (1986). The evolution of multigene families: Human haptoglobin genes. *Ann. Rev. Genet.* 20, 81-108.
27. Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* 302, 575-581.
28. Hood, L., Kronenberg, M., and Hunkapiller, T. (1985). T cell antigen receptors and the immunoglobulin supergene family. *Cell* 40, 225-229.
29. Baltimore, D. (1981). Gene conversion: Some implications for immunoglobulin genes. Minireview. *Cell* 24, 592-594.
30. Egel, R. (1981). Intergenic conversion and reiterated genes. *Nature* 290, 191-192.
31. Flanagan, J.G., Lefranc, M.-P., and Rabbitts, T.H. (1984). Mechanisms of divergence and convergence of the human immunoglobulin $\alpha 1$ and $\alpha 2$ constant region gene sequences. *Cell* 36, 681-688.

-47-

- 5 32. Van der Ploeg, L.H.T. (1991). Antigenic variation in African trypanosomes: genetic recombination and transcriptional control of VSG genes. In Gene Rearrangement. (New York: Oxford University Press), pp 51-98.
- 10 33. Shepherd, G.M. (1985). Are there labeled lines in the olfactory pathway? In Taste, Olfaction and the Central Nervous System (New York: The Rockefeller University Press), pp. 307-321.
- 15 34. Mackay-Sim, A., Shaman, P., and Moulton, D.G. (1982). Topographic coding of olfactory quality: odorant-specific patterns of epithelial responsivity in the Salamander. J. Neurophysiol. 48, 584-596.
- 20 35. Thommesen, G., and Doving, K.B. (1977). Spatial distribution of the EOG in the rat. A variation with odour stimulation. Acta physiol. scand. 99, 270-280.
- 25 36. Sicard, G. (1985). Olfactory discrimination of structurally related molecules: Receptor cell responses to camphoraceous odorant. Brain Res. 326, 203-212.
- 30 37. Thommesen, G. (1978). The spatial distribution of odour induced potentials in the olfactory bulb of char and trout (Salmonidae). Acta Physiol. Scand. 102, 205-217.
38. Doving, K.B., Selset, R., and Thommesen, G. (1980). Olfactory sensitivity to bile acids in salmonid fishes. Acta Physiol. Scand. 108, 123-131.
- 35 39. Stewart, W.B., Kauer, J.S., and Shepherd, G.M. (1979).

-48-

Functional organization of the rat olfactory bulb analyzed by the 2-deoxyglucose method. J. Comp. Neurol. 185, 715-734.

- 5 40. Chirgwin, J.M., Przbyla, A.E., MacDonald, R.J., and Rutter, W.J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18, 5294-5299.
- 10 41. Maniatis, T., Fritsch, E., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).
- 15 42. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).
- 20 43. Julius et al., (1988). Molecular Characterization of a Functional cDNA Encoding a Serotonin 1c Receptor. Science 241, 558-564.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Columbia University in the City of N.Y.,
The Trustees of
- (ii) TITLE OF INVENTION: ODORANT RECEPTORS AND USES THEREOF
- (iii) NUMBER OF SEQUENCES: 36
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: COOPER & DUNHAM
 - (B) STREET: 30 Rockefeller Plaza
 - (C) CITY: New York
 - (D) STATE: New York
 - (E) COUNTRY: U.S.A.
 - (F) ZIP: 10112
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 681,880
 - (B) FILING DATE: 05-APR-1991
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: White, John P.
 - (B) REGISTRATION NUMBER: 28,678
 - (C) REFERENCE/DOCKET NUMBER: 38586
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (212) 977-9550
 - (B) TELEFAX: (212) 664-0525
 - (C) TELEX: (212) 422523 COOP UI

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 954 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: YES
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: rat olfactory epithelium
 - (B) STRAIN: Sprague-Dawley rat
 - (F) TISSUE TYPE: olfactory epithelium
- (vii) IMMEDIATE SOURCE:

SUBSTITUTE SHEET

-50-

(B) CLONE: F12

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGGAATCAG GGAACAGCAC AAGAAGATTT TCAAGTTTTT TTCTTCTTGG ATTTACAGAA	60
AACCCACAAC TTCACTTCCT CATTTTGGCA CTATTCCTGT CCATGTACCT GGTAACAGTG	120
CTTGGGAACC TGCTTATCAT TATGGCCATC ATCACACAGT CTCATTGCA TACACCCATG	180
TACTTTTCC TTGCTAACCT ATCCTTGTG GACATCTGTT TCACCTCCAC CACCATCCCA	240
AAGATGTTGG TAAATATATA CACCCAGAGC AAGAGCATCA CCTATGAAGA CTGTATTAGC	300
CAGATGTGTG TCTTCTTGGT TTTCGCAGAA TTGGGCAACT TTCTCCTGGC TGTGATGGCC	360
TATGACCGAT ATGTGGCTAA CTGTCACCCA CTGTGTTACA CAGTCATTGT GAACCACCGG	420
CTCTGTATCC TGCTGCTTCT GCTGTCCTGG GTTATCAGCA TTTTCCATGC CTTCATAAG	480
AGCTTAATTG TGCTACAGTT GACCTTCTGT GGAGATGTGA AATCCCTCA CTTCTTCTGT	540
GAACCTAATC AGCTGTCCCA ACTCACCTGT TCAGACAACCT TTCCAAGTCA CCTCATAATG	600
AATCTTGATC CTGTTATGTT GGCAGCCATT TCCTTCAGTG GCATCCTTTA CTCTTATTTT	660
AAGATAGTAT CCTCCATACA TTCTATCTCC ACAGTTCAGG GGAAGTACAA GGCATTTTCT	720
ACTTGTCCT CTCACCTTTC CATTGTCTCC TTATTTTATA GTACAGGCCT CCGAGTGTAC	780
GTCAGTTCTG CTGTGGTCCA AAGCTCACAT TCTGCTGCAA GTGCTTCGGT CATGTATACT	840
GTGGTCACCC CCATGCTGAA CCCCTTCATT TATAGTCTAA GGAATAAAGA TGTGAAGAGA	900
GCTCTGGAAG GACTGTTAGA AGGAACTGT AAAGTGCATC ATTGGACTGG ATGA	954

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1002 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: F3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

SUBSTITUTE SHEET

-51-

ATGGACTCAA GCAACAGGAC AAGAGTTTCA GAATTTCTTC TTCTTGGATT TGTAGAAAAC	60
AAAGACCTAC AACCCCTTAT TTATGGTCTT TTTCTCTCTA TGTACCTGGT TACTGTCATT	120
GGAAACATAT CCATTATTGT GGCTATCATT TCAGATCCCT GTCTGCACAC CCCCATGTAT	180
TTCTTCCTCT CTAACCTGTC CTTTGTGGAC ATCTGTTTCA TTTCAACCAC TGTTCCAAAG	240
ATGTTAGTGA ACATCCAGAC CCAAAACAAT GTCATCACCT ATGCAGGATG CATTACCCAG	300
ATATACTTTT TCTTGCTCTT TGTAGAATTG GACAACTTCT TGCTGACTAT CATGGCCTAT	360
GACCGTTAAG TAGCCATCTG TCACCCCATG CACTACACAG TTATCATGAA CTACAAGCTC	420
TGTGGATTTC TGGTTCTGGT ATCTTGGATT GTAAGTGTTT TGCATGCCTT GTTTCAAAGC	480
TTGATGATGT TGGCGCTGCC CTTCTGCACA CATCTGGAAA TCCCACACTA CTTCTGTGAA	540
CCTAATCAGG TGATTCAACT CACCTGTTCT GATGCATTTT TTAATGATCT TGTGATATAT	600
TTTACACTTG TGCTGCTGGC TACTGTTCTT CTTGCTGGCA TCTTCTATTC TTAATTCAAG	660
ATAGTGTCTT CCATATGTGC TATATCGTCA GTTCATGGGA AGTACAAAGC ATTCTCCACC	720
TGTGCATCTC ACCTTTCAGT CGTGTCTTTA TTTTACTGCA CAGGACTAGG AGTGTACCTC	780
AGTTCTGCTG CAAACAACAG CTCACAGGCA AGTGCCACAG CCTCAGTCAT GTACACTGTA	840
GTTACCCCTA TGGTGAACCC TTTTATCTAT AGTCTTAGGA ATAAAGATGT TAAGAGTGTT	900
CTGAAAAAAA CTCTTTGTGA GGAAGTTATA AGGAGTCCAC CTTCCCTACT TCATTTCTTC	960
CTAGTGTTAT GTCATCTCCC TTGTTTTATT TTTTGTTATT AA	1002

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 942 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: F5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGAGCAGCA CCAACCAGTC CAGTGTACC GAGTTCCTCC TCCTGGGACT CTCCAGGCAG	60
CCCCAGCAGC AGCAGCTCCT CTTCTGCTC TTCCTCATCA TGTACCTGGC CACTGTCCTG	120

SUBSTITUTE SHEET

GGAAACCTGC TCATCATCCT GGCTATTGGC ACAGACTCCC GCCTGCACAC CCCCATGTAC 180
TTCTTCCTCA GTAACCTGTC CTTTGTGGAT GTCTGCTTCT CCTCTACCAC TGTCCCTAAA 240
GTTCTGGCCA ACCATATACT TGGGAGTCAG GCCATTTCTT TCTCTGGGTG TCTCACCAG 300
CTGTATTTTC TCGCTGTGTT TGGTAACATG GACAATTTCC TGCTGGCTGT GATGTCCTAT 360
GACCGATTTG TGGCCATATG CCACCCTTTA CACTACACAA CAAAGATGAC CCGTCAGCTC 420
TGTGTCCTGC TTGTTGTGGG GTCATGGGTT GTAGCCAACA TGAATTGTCT GTTGACATA 480
CTGCTCATGG CTGACTCTC CTTCTGTGCA GACAACATGA TCCCCACTT CTTCTGTGAT 540
GGAACTCCCC TCCTGAAACT CTCCTGCTCA GACACACATC TCAATGAGCT GATGATTCTT 600
ACAGAGGGAG CTGTGGTCAT GGTACCCCCA TTTGTCTGCA TCCTCATCTC CTACATCCAC 660
ATCACCTGTG CTGTCCTCAG AGTCTCATCC CCCAGGGGAG GATGGAAATC CTTCTCCACC 720
TGTGGCTCCC ACCTGGCTGT GGTCTGCCTC TTCTATGGCA CCGTCATCGC TGTGTATTTT 780
AACCCATCAT CCTCTCACTT AGCTGGGAGG GACATGGCAG CTGCAGTGAT GTATGCAGTG 840
GTGACCCCAA TGCTGAACCC TTTCATCTAT AGCCTGAGGA ACAGCGACAT GAAAGCAGCT 900
TTAAGGAAAG TGCTGCCCAT GAGATTTCCT TCTAAGCAGT AA 942

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 936 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: YES
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: rat olfactory epithelium
 - (B) STRAIN: Sprague-Dawley rat
 - (F) TISSUE TYPE: olfactory epithelium
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: P6

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATGGCTTGA GTACTGGCCA GAACCTGTCC ACACCAGGAC CATTCATCTT GCTGGGCTTC 60
CCAGGGCCAA GGAGCATGCG CATTGGGCTC TTCCTGCTTT TCCTGGTCAT GTATCTGCTT 120
ACGGTAGTTG GAAACCTAGC CATCATCTCC CTGGTAGGTG CCCACAGATG CCTACAGACA 180
CCCATGTACT TCTTCCTCTG CAACCTCTCC TTCCTGGAGA TCTGGTTCAC CACAGCCTGC 240
GTACCCAAGA CCCTGGCCAC ATTTGCGCCT CGGGGTGGAG TCATTTCCTT GGCTGGCTGT 300

SUBSTITUTE SHEET

GCCACACAGA TGIACCTTGT CTTTCTTTG GGCTGTACOG AGTACTTCCT GCTGGCTGTG	360
ATGGCTTATG ACOGCTACCT GGCCATCTGC CTGCCACTGC GCTATGGTGG CATCATGACT	420
CCTGGGCTGG CGATGCGGTT GGCCCTGGGA TCCTGGCTGT GTGGGTTTTC TGCAATCACA	480
GTTCTGCTA CCCTCATTGC CCGCCTCTCT TTCTGTGGCT CACGTGTCAT CAACCACTTC	540
TTCTGTGACA TTTCGCCCTG GATAGTGCTT TCCTGCACCG ACACGCAGGT GGTGGAAGTG	600
GTGTCCTTTG GCATTGCCTT CTGTGTTATT CTGGGCTCGT GTGGTATCAC ACTAGTCTCC	660
TATGCTTACA TCATCACTAC CATCATCAAG ATTCCCTCTG CCGGGGGCCG GCACCGCGCC	720
TTCTCAACCT GCTCATCCCA TCTCACTGTG GTGCTGATTT GGTATGGCTC CACCATCTTC	780
TTGCATGTGA GGACCTCGGT AGAGAGCTCC TTGGACCTCA CCAAAGCTAT CACAGTGCTC	840
AACACCATTG TCACACCTGT GCTGAACCCT TTCATATATA CTCTGAGGAA CAAGGATGTC	900
AAGGAAGCTC TGCGCAGGAC GGTGAAGGGG AAGTGA	936

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 939 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: I14

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATGACTGGAA ATAACCAAAC TTGATCTTG GAGTTCCTCC TCCTGGGTCT GCCCATCCCA	60
TCAGAGTATC ATCTCCTGTT CTATGCCCTG TTCCTGGCCA TGTACCTCAC CATCATCCTG	120
GGAAACCTGC TAATCATTGT CCTGTTCGA CTGGACTCTC ATCTCCACAT GCCCATGTAC	180
TTGTTTCTCA GCAACTTGTC CTTCTCTGAC CTCTGCTTTT CCTCTGTAC AATGCCCAAA	240
TTGCTTCAGA ACATGCAGAG CCAAGTACCA TCTATATCCT ATACAGGCTG CCTGACACAG	300
CTGTACTTCT TTATGGTTTT TGGAGATATG GAGAGCTTCC TTCTTGTTGGT CATGGCCTAT	360
GACCGCTATG TGGCCATTTG CTTTCCTTTG CGTTACACCA CCATCATGAG CACCAAGTTC	420
TGTGCTTCAC TAGTGCTACT TCTGTGGATG CTGACGATGA CCCATGCCCT GCTGCATACC	480

SUBSTITUTE SHEET

-54-

CTACTCATTG CTAGATTGTC TTTTGTGAG AAGAATGTGA TTCTTCACTT TTTCTGTGAC 540
ATTTCTGCTC TTCTGAAGTT GTCCTGCTCA GACATTTATG TTAATGAGCT GATGATATAT 600
ATCTTGGGTG GACTCATCAT TATTATCCCA TTCCTATTAA TTGTTATGTC CTATGTTAGA 660
ATTTTCTTCT CCATTTTGAA GTTCCATCT ATTCAAGACA TCTACAAGGT ATTCTCAACC 720
TGTGGTTCCC ATCTGTCTGT GGTGACCTTG TTTTATGGGA CAATTTTGG TATCTACTTA 780
TGTCCATCAG GTAATAATTC TACTGTGAAG GAGATTGCCA TGGCTATGAT GTACACAGTG 840
GTGACTCCCA TGCTGAATCC CTTCACTAC AGCCTGAGGA ACAGAGACAT GAAAAGGGCC 900
CTAATAAGAG TTATCTGCAC TAAGAAAATC TCTCTGTAA 939

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 945 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: I15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

ATGACAGAAG AGAACCAAC TGTGATCTCC CAGTTCCTTC TCCTTTTCCT GCCCATCCCC 60
TCAGAGCACC AGCAGTGTT CTACGCCCTG TTCCTGTCCA TGTACCTCAC CACTGTCCTG 120
GGGAACCTCA TCATCATCAT CCTCATTAC CTGGACTCCC ATCTCCACAC ACCCATGTAC 180
TTGTTTCTCA GCAACTTGTG CTTCTCTGAT CTCTGCTTTT CCTCTGTTAC GATGCCCAAG 240
TTGTTGCAGA ACATGCAGAG CCAAGTTCCA TCCATCCCCT TTGCAGGCTG CCTGACACAA 300
TTATACTTTT ACCTGTATTT TGCAGACCTT GAGAGCTTCC TGCTTGTGGC CATGGCCTAT 360
GACCGCTATG TGGCCATCTG CTTCCCCCTT CATTACATGA GCATCATGAG CCCCAAGCTC 420
TGTGTGAGTC TGGTGGTGCT GTCCTGGGTG CTGACCACCT TCCATGCCAT GCTGCACACC 480
CTGCTCATGG CCAGATTGTC ATTCTGTGCG GACAATATGA TCCCCCACTT TTTCTGTGAT 540
ATATCTCCTT TATTGAACT GTCCTGCTCT GACACGCATG TTAATGAGTT GGTGATATTT 600

SUBSTITUTE SHEET

GTCATGGGAG GGCTTGTTAT TGTCATTCCA TTTGTGCTCA TCATTGTATC TTATGCACGA	660
GTTGTGCGCT CCATTCTTAA AGTCCCTTCT GTCCGAGGCA TCCACAAGAT CTTCTCCACC	720
TGCGGCTCCC ATCTGTCTGT GGTGTCAC TGTCATGGGA CAATCATTGG TCTCTACTTA	780
TGTCCGTCAG CTAATAACTC TACTGTGAAG GAGACTGTCA TGGCCATGAT GTACACAGTG	840
GTGACCCCCA TGCTGAACCC CTTTCATCTAC AGCCTGAGGA ACAGAGACAT GAAAGAGGCA	900
CTGATAAGAG TCCTTTGTAA AAAGAAAATT ACCTTCTGTC TATGA	945

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 933 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: I3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATGAACAATC AAACCTTCAT CACCCAATTC CTTCTCCTGG GACTGCCCAT CCCTGAAGAA	60
CATCAGCACC TGTTCTATGC CTTGTTCTTG GTCATGTACC TCACCACCAT CTTGGGAAAC	120
TTGCTAATCA TTGTACTTGT TCAACTGGAC TCCCAGCTCC ACACACCTAT GTATTTGTTT	180
CTCAGCAATT TGTCTTTCTC TGATCTATGT TTTTCCTCTG TCACAATGCC CAAGCTGCTG	240
CAGAACATGA GGAGCCAGGA CACATCCATT CCCTATGGAG GCTGCCTGGC ACAAACATAC	300
TTCTTTATGG TTTTGGAGA TATGGAGAGT TTCCTTCTTG TGGCCATGGC CTATGACCGC	360
TATGTGGCCA TCTGCTTCCC TCTGCATTAC ACCAGCATCA TGAGCCCCAA GCTCTGTACT	420
TGTCTAGTGC TGTTATTGTG GATGCTGACG ACATCCCATG CCATGATGCA CACACTGCTT	480
GCAGCAAGAT TGTCTTTTGG TGAGAACAAT GTGGTCCTCA ACTTCTTCTG TGACCTATTT	540
GTTCTCCTAA AGCTGGCCTG CTCAGACACT TATATTAATG AGTTGATGAT ATTTATCATG	600
AGTACACTCC TCATTATTAT TCCATTCTTC CTCATTGTTA TGTCTATGC AAGGATCATA	660
TCCTCTATTC TTAAGGTTCC ATCTACCCAA GGCATCTGCA AGGTCTTCTC TACCTGTGGT	720

TCCCATCTGT CTGTAGTATC ACTGTTCTAT GGGACAATTA TTGGTCTCTA CTTATGTCCA 780
GCAGGTAATA ATTCCACTGT AAAAGAGATG GTCATGGCCA TGATGTACAC TGTGGTGACC 840
CCCATGCTGA ATCCCTTCAT CTACAGCCTA AGGAATAGAG ATATGAAGAG GGCCCTAATA 900
AGAGTTATCT GTAGTATGAA AATCACTCTG TAA 933

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 984 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: I7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATGGAGCGAA GGAACCACAG TGGGAGAGTG AGTGAATTTG TGTGCTGGG TTTCCCAGCT 60
CCTGCCCCAC TGGGAGTACT ACTATTTTTC CTTTCTCTTC TGGACTATGT GTTGGTGTTG 120
ACTGAAAACA TGCTCATCAT TATAGCAATT AGGAACCACC CAACCCTCCA CAAACCCATG 180
TATTTTTTCT TGGCTAATAT GTCATTTCTG GAGATTTGGT ATGTCACTGT TACGATTCCT 240
AAGATGCTCG CTGGCTTCAT TGGTTCCAAG GAGAACCATG GACAGCTGAT CTCCTTTGAG 300
GCATGCATGA CACAACCTCA CTTTTTCCTG GGCTTGGGTT GCACAGAGTG TGTCCCTTCTT 360
GCTGTGATGG CCTATGACCG CTATGTGGCT ATCTGTCATC CACTCCACTA CCCCGTCATT 420
GTCAGTAGCC GGCTATGTGT GCAGATGGCA GCTGGATCCT GGGCTGGAGG TTTTGGTATC 480
TCCATGGTTA AAGTTTTCCT TATTTCTCGC CTGTCTTACT GTGGCCCCAA CACCATCAAC 540
CACTTTTTCT GTGATGTGTC TCCATTGCTC AACCTGTCAT GCACTGACAT GTCCACAGCA 600
GAGCTTACAG ACTTTGTCCT GGCCATTTTT ATTCTGCTGG GACCGCTCTC TGTCACTGGG 660
GCATCCTACA TGGCCATCAC AGGTGCTGTG ATGCGCATCC CCTCAGCTGC TGGCCGCCAT 720
AAAGCCTTTT CAACCTGTGC CTCCCACCTC ACTGTTGTGA TCATCTTCTA TGCAGCCAGT 780
ATTTTCATCT ATGCCAGGCC TAAGGCACTC TCAGCTTTTG ACACCAACAA GCTGGTCTCT 840
GTACTCTACG CTGTCATTGT ACCGTTGTTT AATCCCATCA TCTACTGCTT GCGCAACCAA 900

SUBSTITUTE SHEET

-57-

GATGTCAAAA GAGCGCTACG TCGCAGCTG CACCTGGCCC AGGACCAGGA GGCCAATACC 960
AACAAAGGCA GCAAAATTGG TTAG 984

(2) INFORMATION FOR SEQ ID NO:9:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 939 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: 18

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATGAACAACA AAACGTGCAT CACCCATTTC CTCCTCCTGG GATTGCCCAT CCCCCCAGAG 60
CACCAGCAAC TGTTCCTTGC CCGTTCCTG ATCATGTACC TCACCACCTT TCTGGGAAAC 120
CTGCTAATTG TTGTCCTTGT TCAACTGGAC TCTCATCTCC ACACACCCAT GTACTTGTTT 180
CTCAGCAACT TGTCTTCTC TGATCTCTGC TTTTCCTCTG TTACAATGCT GAAATTGCTG 240
CAAAATATAC AGAGCCAAGT ACCATCTATA TCCTATGCAG GATGCCTGAC ACAGATATTC 300
TTCTTTTTGT TGTTCGGCTA CCGTGGGAAT TTCCTTCTTG TAGCCATGGC CTATGACCGC 360
TATGTGGCCA TCTGCTTCCC TCTGCATTAT ACCAACATCA TGAGCCATAA GCTCTGTACT 420
TGTCTCCTGC TGGTATTTTG GATAATGACA TCATCTCATG CCATGATGCA CACCCTGCTT 480
GCAGCAAGAT TGTCTTTTTG TGAGAACAAT GTACTCCTCA ACTTTTCTG TGACCTGTTT 540
GTTCTCCTAA AGTTGGCCTG CTCAGACACT TATGTTAATG AGTTGATGAT ACATATCATG 600
GGCGTGATCA TCATTGTTAT TCCATTGCTG CTCATTGTTA TATCCTATGC CAAGATCATC 660
TCCTCCATTC TTAAGGTTCC ATCTACTCAA AGCATTCA CA AGGTCTTCTC CACTTGTTGGT 720
TCTCATCTCT CTGTGGTGTG TCTGTTCTAC GGGACAATTA TTGGTCTCTA TTTATGTCCA 780
TCAGGTGATA ATTTTAGTCT AAAGGGGTCT GCCATGGCTA TGATGTACAC AGTGGTAACT 840
CCAATGCTGA ACCCGTTCAT CTACAGCCTA AGAAACAGAG ACATGAAGCA GGCCCTAATA 900
AGAGTTACCT GTAGCAAGAA AATCTCTCTG CCATGCTAG 939

(2) INFORMATION FOR SEQ ID NO:10:

SUBSTITUTE SHEET

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 945 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: Sprague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: 19

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATGACTAGAA GAAACCAAAC TGCCATCTCT CAGTTCTTCC TTCTGGGCCT GCCATTCCCC	60
CCAGAGTACC AACACCTGTT CTATGCCCTG TTCCTGGCCA TGTACCTCAC CACTCTCCTG	120
GGGAACCTCA TCATCATCAT CCTCATTCTA CTGGACTCCC ATCTCCACAC ACCCATGTAC	180
TTGTTTCTCA GCAATTTATC CTTTGCCGAC CTCTGTTTTT CCTCTGTCAC AATGCCCAAG	240
TTGTTGCAGA ACATGCAGAG CCAAGTTCCA TCCATCCCCCT ATGCAGGGTG CCTGGCACAG	300
ATATACTTCT TTCTGTTTTT TGGAGACCTT GGAAACTTCC TGCTTGTGGC CATGGCCTAT	360
GACCGCTATG TGGCCATCTG CTTCCCCCTT CATTACATGA GCATCATGAG CCCCAGCTC	420
TGTGTGAGTC TGGTGGTGCT GTCCTGGGTG CTGACTACCT TCCATGCCAT GCTGCACACC	480
CTGCTCATGG CCAGATTGTC ATTCTGTGAG GACAGTGTGA TCCCTCACTA TTTCTGTGAT	540
ATGTCTACTC TGCTGAAAGT GGCTTGTCTT GACACCCATG ATAATGAATT AGCAATATTT	600
ATCTTAGGGG GCCCTATAGT TGTACTACCT TTCCTTCTCA TCATTGTTTC TTATGCAAGA	660
ATTGTTTCCT CCATCTTCAA GGTCCCTTCT TCTCAAAGCA TCCATAAAGC CTTCTCCACC	720
TGTGGCTCCC ACCTGTCTGT GGTGTCACTG TTCTATGGGA CAGTCATTGG TCTCTACTTA	780
TGTCCTTCAG CTAATAACTC CACTGTGAAG GAGACTGTCA TGTCTTTGAT GTACACAATG	840
GTGACACCCA TGCTGAACCC CTTCATCTAC AGCCTAAGAA ACAGAGACAT AAAAGATGCA	900
TTAGAAAAAA TAATGTGCAA AAAGCAAATT CCCTCCTTTC TATGA	945

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 645 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

SUBSTITUTE SHEET

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapien

(vii) IMMEDIATE SOURCE:

(B) CLONE: H5

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..645

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATC TGT TTT GTG TCT ACC ACT GTC CCA AAG CAG CTG GTG AAC ATC CAG	48
Ile Cys Phe Val Ser Thr Thr Val Pro Lys Gln Leu Val Asn Ile Gln	
1 5 10 15	
ACA CAG AGC AGA GTC ATC ACC TAT GCA GAC TGC ATC ACC CAG ATG TGC	96
Thr Gln Ser Arg Val Ile Thr Tyr Ala Asp Cys Ile Thr Gln Met Cys	
20 25 30	
TTT TTT ATA CTC TTT GTA GTG TTG GAC AGC TTA CTC CTG ACT GTG ATG	144
Phe Phe Ile Leu Phe Val Val Leu Asp Ser Leu Leu Leu Thr Val Met	
35 40 45	
GCC TAT GAC CGG TTT GTG GCC ATC TGT CAC CCC CTG CAC TAC ACA GTC	192
Ala Tyr Asp Arg Phe Val Ala Ile Cys His Pro Leu His Tyr Thr Val	
50 55 60	
ATT ATG AGC TCC TGG CTC TGT GGA CTG CTG GTT CTG GTG TCC TGG ATC	240
Ile Met Ser Ser Trp Leu Cys Gly Leu Leu Val Leu Val Ser Trp Ile	
65 70 75 80	
GTG AGC ATC CTA TAT TCT CTG TTA CAA AGC ATA ATG GCA TTG CAG CTG	288
Val Ser Ile Leu Tyr Ser Leu Leu Gln Ser Ile Met Ala Leu Gln Leu	
85 90 95	
TCC TTC TGT ACA GAA CTG AAA ATC CCT CAA TTT TTC TGT GAA CTT AAT	336
Ser Phe Cys Thr Glu Leu Lys Ile Pro Gln Phe Phe Cys Glu Leu Asn	
100 105 110	
CAG GTC ATC CAC CTT GCC TGT TCC GAC ACT TTT ATT AAT GAC ATG ATG	384
Gln Val Ile His Leu Ala Cys Ser Asp Thr Phe Ile Asn Asp Met Met	
115 120 125	
ATG AAT TTT ACA AGT GTG CTG CTG GGT GGG GGA TGC CTC GCT GGA ATA	432
Met Asn Phe Thr Ser Val Leu Leu Gly Gly Gly Cys Leu Ala Gly Ile	
130 135 140	
TTT TAC TNN TAC TTT AAG ATA CTT TGT TGC ATA TGT TCG ATC TCA TCA	480
Phe Tyr Xaa Tyr Phe Lys Ile Leu Cys Cys Ile Cys Ser Ile Ser Ser	
145 150 155 160	
GCT CAG GGG ATG AAT AAA GCA CTT TCC ACC TGT GCA TCT CAC CTC TCA	528
Ala Gln Gly Met Asn Lys Ala Leu Ser Thr Cys Ala Ser His Leu Ser	
165 170 175	

SUBSTITUTE SHEET

-60-

GTT GTC TCC TTA TTT TAT TGT ACA GGC GTA GGT GTG TAC CTT AGT TCT	576
Val Val Ser Leu Phe Tyr Cys Thr Gly Val Gly Val Tyr Leu Ser Ser	
180 185 190	
GCT GCA ACC CAT AAC TCA CTC TCA AAT GCT GCA GCC TCG GTG ATG TAC	624
Ala Ala Thr His Asn Ser Leu Ser Asn Ala Ala Ala Ser Val Met Tyr	
195 200 205	
ACT GTG GTC ACC TCC ATG CTG	645
Thr Val Val Thr Ser Met Leu	
210 215	

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 215 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Ile Cys Phe Val Ser Thr Thr Val Pro Lys Gln Leu Val Asn Ile Gln	
1 5 10 15	
Thr Gln Ser Arg Val Ile Thr Tyr Ala Asp Cys Ile Thr Gln Met Cys	
20 25 30	
Phe Phe Ile Leu Phe Val Val Leu Asp Ser Leu Leu Leu Thr Val Met	
35 40 45	
Ala Tyr Asp Arg Phe Val Ala Ile Cys His Pro Leu His Tyr Thr Val	
50 55 60	
Ile Met Ser Ser Trp Leu Cys Gly Leu Leu Val Leu Val Ser Trp Ile	
65 70 75 80	
Val Ser Ile Leu Tyr Ser Leu Leu Gln Ser Ile Met Ala Leu Gln Leu	
85 90 95	
Ser Phe Cys Thr Glu Leu Lys Ile Pro Gln Phe Phe Cys Glu Leu Asn	
100 105 110	
Gln Val Ile His Leu Ala Cys Ser Asp Thr Phe Ile Asn Asp Met Met	
115 120 125	
Met Asn Phe Thr Ser Val Leu Leu Gly Gly Gly Cys Leu Ala Gly Ile	
130 135 140	
Phe Tyr Xaa Tyr Phe Lys Ile Leu Cys Cys Ile Cys Ser Ile Ser Ser	
145 150 155 160	
Ala Gln Gly Met Asn Lys Ala Leu Ser Thr Cys Ala Ser His Leu Ser	
165 170 175	
Val Val Ser Leu Phe Tyr Cys Thr Gly Val Gly Val Tyr Leu Ser Ser	
180 185 190	
Ala Ala Thr His Asn Ser Leu Ser Asn Ala Ala Ala Ser Val Met Tyr	
195 200 205	

SUBSTITUTE SHEET

-61-

Thr Val Val Thr Ser Met Leu
210 215

(2) INFORMATION FOR SEQ ID NO:13:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 640 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
(B) STRAIN: Sprague-Dawley rat
(F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

(B) CLONE: J1

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 2..640

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

C ATC TGC TTT ACT TCT GCT AGC ATC CCA AAG ATG CTA GTG AAT ATA	46
Ile Cys Phe Thr Ser Ala Ser Ile Pro Lys Met Leu Val Asn Ile	
1 5 10 15	
CAG ACG AAG AAC AAG GTG ATC ACC TAT GAA GGC TGC ATC TCC CAA GTA	94
Gln Thr Lys Asn Lys Val Ile Thr Tyr Glu Gly Cys Ile Ser Gln Val	
20 25 30	
TAC TTT TCA TAC TCT TTG GAG TTT TGG ACA ACT TTC TTC TCG ACT GTG	142
Tyr Phe Ser Tyr Ser Leu Glu Phe Trp Thr Thr Phe Phe Ser Thr Val	
35 40 45	
ATG GCC TAT GAC CGA TAT GTG GCC ATC TGT CAC CCA TCT NAC TAC ACA	190
Met Ala Tyr Asp Arg Tyr Val Ala Ile Cys His Pro Ser Xaa Tyr Thr	
50 55 60	
GGT CAT CAT GAA CCN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN	238
Gly His His Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
65 70 75	
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN	286
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
80 85 90 95	
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN	334
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	
100 105 110	
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN	382
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa	

SUBSTITUTE SHEET

-62-

115	120	125	
NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NTT			430
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa			
130	135	140	
TAT TCT TAC TCT AAG ATA GTT TCC TCC ATA CGA GAA ATC TCA TCA TCA			478
Tyr Ser Tyr Ser Lys Ile Val Ser Ser Ile Arg Glu Ile Ser Ser Ser			
145	150	155	
CAG GGA AAG TAC AAG NNA TTC TCC ACC TGT GCA TCC CAC CTC TCA GTT			526
Gln Gly Lys Tyr Lys Xaa Phe Ser Thr Cys Ala Ser His Leu Ser Val			
160	165	170	175
GTT TCA TTA TTC TAT TCT ACA CTT TTG GGT GTG TAC CTT AGT TCT TCT			574
Val Ser Leu Phe Tyr Ser Thr Leu Leu Gly Val Tyr Leu Ser Ser Ser			
180	185	190	
TTT ACC CAA AAC TCA CAC TCA ACT GCA CGG GCA TCT GTT ATG TAC AGT			622
Phe Thr Gln Asn Ser His Ser Thr Ala Arg Ala Ser Val Met Tyr Ser			
195	200	205	
GTG GTC ACC CCC ATG TTG			640
Val Val Thr Pro Met Leu			
210			

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 213 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Ile Cys Phe Thr Ser Ala Ser Ile Pro Lys Met Leu Val Asn Ile Gln			
1	5	10	15
Thr Lys Asn Lys Val Ile Thr Tyr Glu Gly Cys Ile Ser Gln Val Tyr			
20	25	30	
Phe Ser Tyr Ser Leu Glu Phe Trp Thr Thr Phe Phe Ser Thr Val Met			
35	40	45	
Ala Tyr Asp Arg Tyr Val Ala Ile Cys His Pro Ser Xaa Tyr Thr Gly			
50	55	60	
His His Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa			
65	70	75	80
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa			
85	90	95	
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa			
100	105	110	
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa			
115	120	125	

SUBSTITUTE SHEET

-63-

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Tyr
 130 135 140
 Ser Tyr Ser Lys Ile Val Ser Ser Ile Arg Glu Ile Ser Ser Ser Gln
 145 150 155 160
 Gly Lys Tyr Lys Xaa Phe Ser Thr Cys Ala Ser His Leu Ser Val Val
 165 170 175
 Ser Leu Phe Tyr Ser Thr Leu Leu Gly Val Tyr Leu Ser Ser Ser Phe
 180 185 190
 Thr Gln Asn Ser His Ser Thr Ala Arg Ala Ser Val Met Tyr Ser Val
 195 200 205
 Val Thr Pro Met Leu
 210

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 636 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: erpague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: J2

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..636

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

ACC	TCC	ACC	ACC	ATC	CCA	AAG	ATG	CTG	GTA	AAT	ATA	CAC	ACC	CAG	AGC	48
Thr	Ser	Thr	Thr	Ile	Pro	Lys	Met	Leu	Val	Asn	Ile	His	Thr	Gln	Ser	
1				5				10						15		
AAT	ACT	ATC	ACC	TAT	GAA	GAC	TGT	ATT	TCC	CAG	ATG	TTT	GTA	CTC	TTG	96
Asn	Thr	Ile	Thr	Tyr	Glu	Asp	Cys	Ile	Ser	Gln	Met	Phe	Val	Leu	Leu	
			20				25						30			
GTT	TTT	GGA	GAA	CTG	GAC	AAC	TTT	CTC	CTG	GCT	GTG	ATG	GCC	TAT	GAT	144
Val	Phe	Gly	Glu	Leu	Asp	Asn	Phe	Leu	Leu	Ala	Val	Met	Ala	Tyr	Asp	
		35				40					45					
CGA	TAT	GTG	GCT	ATC	TGT	CAC	CCA	CTG	TAT	TAC	ACA	GTC	ATT	GTG	AAC	192
Arg	Tyr	Val	Ala	Ile	Cys	His	Pro	Leu	Tyr	Tyr	Thr	Val	Ile	Val	Asn	
		50				55					60					

SUBSTITUTE SHEET

-64-

CAC CGA CTC TGT ATC CTG CTG CTT CTG CTG TCC TGG GTT GTC AGC ATT	240
His Arg Leu Cys Ile Leu Leu Leu Leu Leu Ser Trp Val Val Ser Ile	
65 70 75 80	
TTA CAT GCC TTC TTA CAG AGC TTA ATT GTA CTA CAG TTG ACC TTC TGT	288
Leu His Ala Phe Leu Gln Ser Leu Ile Val Leu Gln Leu Thr Phe Cys	
85 90 95	
GGA GAT GTG AAA ATC CCT CAC TTC TTC TGT GAG CTC AAT CAG CTG TCC	336
Gly Asp Val Lys Ile Pro His Phe Phe Cys Glu Leu Asn Gln Leu Ser	
100 105 110	
CAA CTC ACA TGT TCA GAC AAC TTT CCA AGT CAC CTC ACA ATG CAT CTT	384
Gln Leu Thr Cys Ser Asp Asn Phe Pro Ser His Leu Thr Met His Leu	
115 120 125	
GTA CCT GTT ATA TTT GCA GCT ATT TCC CTC AGT GGT ATC CTT TAC TCT	432
Val Pro Val Ile Phe Ala Ala Ile Ser Leu Ser Gly Ile Leu Tyr Ser	
130 135 140	
TAT TTC AAG ATA GTG TCT TCC ATA CGT TCT ATG TCC TCA GTT CAA GCG	480
Tyr Phe Lys Ile Val Ser Ser Ile Arg Ser Met Ser Ser Val Gln Gly	
145 150 155 160	
AAG TAC AAG GCA TTT TCT ACA TGT GCC TCT CAC CTT TCC ATT GTC TCC	528
Lys Tyr Lys Ala Phe Ser Thr Cys Ala Ser His Leu Ser Ile Val Ser	
165 170 175	
TTA TTT TAT AGT ACA GGC CTC GCG GTG TAC GTC AGT TCT GCT GTG ATC	576
Leu Phe Tyr Ser Thr Gly Leu Gly Val Tyr Val Ser Ser Ala Val Ile	
180 185 190	
CGA AGC TCA CAC TCC TCT GCA AGT GCT TCG GTC ATG TAT ACT GTG GTC	624
Arg Ser Ser His Ser Ser Ala Ser Ala Ser Val Met Tyr Thr Val Val	
195 200 205	
ACC CCC ATG TTG	636
Thr Pro Met Leu	
210	

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 212 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Thr Ser Thr Thr Ile Pro Lys Met Leu Val Asn Ile His Thr Gln Ser	
1 5 10 15	
Asn Thr Ile Thr Tyr Glu Asp Cys Ile Ser Gln Met Phe Val Leu Leu	
20 25 30	
Val Phe Gly Glu Leu Asp Asn Phe Leu Leu Ala Val Met Ala Tyr Asp	
35 40 45	
Arg Tyr Val Ala Ile Cys His Pro Leu Tyr Tyr Thr Val Ile Val Asn	

SUBSTITUTE SHEET

-65-

50		55		60
His Arg Leu Cys Ile Leu Leu Leu Leu Leu Ser Trp Val Val Ser Ile				
65		70		75
				80
Leu His Ala Phe Leu Gln Ser Leu Ile Val Leu Gln Leu Thr Phe Cys				
	85		90	95
Gly Asp Val Lys Ile Pro His Phe Phe Cys Glu Leu Asn Gln Leu Ser				
	100		105	110
Gln Leu Thr Cys Ser Asp Asn Phe Pro Ser His Leu Thr Met His Leu				
	115		120	125
Val Pro Val Ile Phe Ala Ala Ile Ser Leu Ser Gly Ile Leu Tyr Ser				
	130		135	140
Tyr Phe Lys Ile Val Ser Ser Ile Arg Ser Met Ser Ser Val Gln Gly				
	145		150	155
				160
Lys Tyr Lys Ala Phe Ser Thr Cys Ala Ser His Leu Ser Ile Val Ser				
	165		170	175
Leu Phe Tyr Ser Thr Gly Leu Gly Val Tyr Val Ser Ser Ala Val Ile				
	180		185	190
Arg Ser Ser His Ser Ser Ala Ser Ala Ser Val Met Tyr Thr Val Val				
	195		200	205
Thr Pro Met Leu				
210				

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 646 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium
- (B) STRAIN: erpague-Dawley rat
- (F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

- (B) CLONE: J4

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..646

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

C ATA GGC TAT TCA TCT TCT GTC ACA CCC AAT ATG CTT GTC AAC TTC

46

SUBSTITUTE SHEET

-66-

Ile	Gly	Tyr	Ser	Ser	Ser	Val	Thr	Pro	Asn	Met	Leu	Val	Asn	Phe	
1				5				10					15		
CTT	ATA	AAG	CAA	AAT	ACC	ATC	TCA	TAC	CTT	GGA	TGT	TCT	ATA	CAG	TTT
Leu	Ile	Lys	Gln	Asn	Thr	Ile	Ser	Tyr	Leu	Gly	Cys	Ser	Ile	Gln	Phe
			20					25					30		
GGC	TCA	GCT	GCT	TTG	TTT	GGA	GGT	CTT	GAA	TGC	TTC	CTT	CTG	GCT	GCC
Gly	Ser	Ala	Ala	Leu	Phe	Gly	Gly	Leu	Glu	Cys	Phe	Leu	Leu	Ala	Ala
		35				40						45			
ATG	GCG	TAT	GAT	CGT	TTT	GTA	GCA	ATC	TGC	AAC	CCA	CTG	CTT	TAT	TCA
Met	Ala	Tyr	Asp	Arg	Phe	Val	Ala	Ile	Cys	Asn	Pro	Leu	Leu	Tyr	Ser
		50				55					60				
ACG	AAA	ATG	TCC	ACA	CAA	GTC	TGT	GTC	CAG	TTG	GTT	GTG	GGA	TCT	TAT
Thr	Lys	Met	Ser	Thr	Gln	Val	Cys	Val	Gln	Leu	Val	Val	Gly	Ser	Tyr
	65				70			75							
ATA	GGG	GGA	TTT	CTT	AAT	GCC	TCC	TCT	TTT	ACC	CTT	TCC	TTT	TTT	TCC
Ile	Gly	Gly	Phe	Leu	Asn	Ala	Ser	Ser	Phe	Thr	Leu	Ser	Phe	Phe	Ser
80					85			90							95
TTG	TCC	TTC	TGT	GGA	CCA	AAT	AGA	ATC	AAT	CAC	TTT	TAC	TGT	GAT	TTT
Leu	Ser	Phe	Cys	Gly	Pro	Asn	Arg	Ile	Asn	His	Phe	Tyr	Cys	Asp	Phe
			100					105					110		
GCT	CCG	TTA	GTA	GAA	CTT	TCT	TGC	TCT	GAT	GTC	AGT	GTT	CCT	GAT	GCT
Ala	Pro	Leu	Val	Glu	Leu	Ser	Cys	Ser	Asp	Val	Ser	Val	Pro	Asp	Ala
		115				120						125			
GTT	ACC	TCA	TTT	TCT	GCT	GCC	TCA	GTT	ACT	ATG	CTC	ACA	GTG	TTT	ATC
Val	Thr	Ser	Phe	Ser	Ala	Ala	Ser	Val	Thr	Met	Leu	Thr	Val	Phe	Ile
		130				135					140				
ATA	GCC	ATC	TCC	TAT	ACC	TAT	ATC	CTC	ATC	ACC	ATC	CTG	AAG	ATG	CGT
Ile	Ala	Ile	Ser	Tyr	Thr	Tyr	Ile	Leu	Ile	Thr	Ile	Leu	Lys	Met	Arg
	145					150				155					
TCC	ACT	GAG	GGT	CGA	CAG	AAA	GCA	TTC	TCT	ACC	TGC	ACT	TCC	CAC	CTC
Ser	Thr	Glu	Gly	Arg	Gln	Lys	Ala	Phe	Ser	Thr	Cys	Thr	Ser	His	Leu
160					165				170						175
ACT	GCA	GTC	ACT	CTG	TGC	TAT	GGA	ACC	ATC	ACA	TTC	ATC	TAT	GTG	ATG
Thr	Ala	Val	Thr	Leu	Cys	Tyr	Gly	Thr	Ile	Thr	Phe	Ile	Tyr	Val	Met
			180					185					190		
CCC	AAG	TCC	AGC	TAC	TCC	ACA	GAC	CAG	AAC	AAG	GTG	GTG	TCT	GTG	TTT
Pro	Lys	Ser	Ser	Tyr	Ser	Thr	Asp	Gln	Asn	Lys	Val	Val	Ser	Val	Phe
		195						200					205		
TAT	ATG	GTG	GTG	ATC	CCC	ATG	TTG								
Tyr	Met	Val	Val	Ile	Pro	Met	Leu								
		210					215								

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 215 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

SUBSTITUTE SHEET

-67-

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```

Ile Gly Tyr Ser Ser Ser Val Thr Pro Asn Met Leu Val Asn Phe Leu
 1           5           10           15
Ile Lys Gln Asn Thr Ile Ser Tyr Leu Gly Cys Ser Ile Gln Phe Gly
 20           25           30
Ser Ala Ala Leu Phe Gly Gly Leu Glu Cys Phe Leu Leu Ala Ala Met
 35           40           45
Ala Tyr Asp Arg Phe Val Ala Ile Cys Asn Pro Leu Leu Tyr Ser Thr
 50           55           60
Lys Met Ser Thr Gln Val Cys Val Gln Leu Val Val Gly Ser Tyr Ile
 65           70           75           80
Gly Gly Phe Leu Asn Ala Ser Ser Phe Thr Leu Ser Phe Phe Ser Leu
 85           90           95
Ser Phe Cys Gly Pro Asn Arg Ile Asn His Phe Tyr Cys Asp Phe Ala
 100          105          110
Pro Leu Val Glu Leu Ser Cys Ser Asp Val Ser Val Pro Asp Ala Val
 115          120          125
Thr Ser Phe Ser Ala Ala Ser Val Thr Met Leu Thr Val Phe Ile Ile
 130          135          140
Ala Ile Ser Tyr Thr Tyr Ile Leu Ile Thr Ile Leu Lys Met Arg Ser
 145          150          155          160
Thr Glu Gly Arg Gln Lys Ala Phe Ser Thr Cys Thr Ser His Leu Thr
 165          170          175
Ala Val Thr Leu Cys Tyr Gly Thr Ile Thr Phe Ile Tyr Val Met Pro
 180          185          190
Lys Ser Ser Tyr Ser Thr Asp Gln Asn Lys Val Val Ser Val Phe Tyr
 195          200          205
Met Val Val Ile Pro Met Leu
 210          215

```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 481 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: rat olfactory epithelium

SUBSTITUTE SHEET

-68-

(B) STRAIN: Sprague-Dawley rat
(F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:
(B) CLONE: J7

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..481

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

C ATC TGC AAG CCC CTG CAC TAC ACC ACC ATC ATG AAT AAC CGA GTG Ile Cys Lys Pro Leu His Tyr Thr Thr Ile Met Asn Asn Arg Val 1 5 10 15	46
TGC ACA GTT CTA GTC CTC TCC TGT TGG TTT GCT GGC CTG TTG ATC ATC Cys Thr Val Leu Val Leu Ser Cys Trp Phe Ala Gly Leu Leu Ile Ile 20 25 30	94
CTC CCA CCT CTT GGT CAT GGC CTC CAG CTG GAG TTC TGT GAC TCC AAT Leu Pro Pro Leu Gly His Gly Leu Gln Leu Glu Phe Cys Asp Ser Asn 35 40 45	142
GTG ATT GAT CAT TTT GGC TGT GAT GCC TCT CCA ATT CTG CAG ATA ACC Val Ile Asp His Phe Gly Cys Asp Ala Ser Pro Ile Leu Gln Ile Thr 50 55 60	190
TGC TCA GAC ACG GTA TTT ATA GAG AAA ATT GTC TTG GCT TTT GCC ATA Cys Ser Asp Thr Val Phe Ile Glu Lys Ile Val Leu Ala Phe Ala Ile 65 70 75	238
CTG ACA CTC ATC ATT ACT CTG GTA TGT GTT GTT CTC TCC TAC ACA TAC Leu Thr Leu Ile Ile Thr Leu Val Cys Val Val Leu Ser Tyr Thr Tyr 80 85 90 95	286
ATC ATC AAG ACC ATT TTA AAG TTT CCT TCT GCT CAA CAA AGA AAA AAG Ile Ile Lys Thr Ile Leu Lys Phe Pro Ser Ala Gln Gln Arg Lys Lys 100 105 110	334
GCC TTT TCT ACA TGT TCT TCC CAC ATG ATT GTG GTT TCC ATC ACC TAT Ala Phe Ser Thr Cys Ser Ser His Met Ile Val Val Ser Ile Thr Tyr 115 120 125	382
GGG AGC TGT ATT TTC ATC TAC ATC AAA CCT TCA GCG AAG GAA GGG GTA Gly Ser Cys Ile Phe Ile Tyr Ile Lys Pro Ser Ala Lys Glu Gly Val 130 135 140	430
GCC ATC AAT AAG GTT GTA TCT GTG CTC ACA ACA TCA GTC GCC CCT TTG Ala Ile Asn Lys Val Val Ser Val Leu Thr Thr Ser Val Ala Pro Leu 145 150 155	478
CTC Leu 160	481

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 160 amino acids

SUBSTITUTE SHEET

-69-

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```

Ile Cys Lys Pro Leu His Tyr Thr Thr Ile Met Asn Asn Arg Val Cys
 1           5           10           15
Thr Val Leu Val Leu Ser Cys Trp Phe Ala Gly Leu Leu Ile Ile Leu
 20           25           30
Pro Pro Leu Gly His Gly Leu Gln Leu Glu Phe Cys Asp Ser Asn Val
 35           40           45
Ile Asp His Phe Gly Cys Asp Ala Ser Pro Ile Leu Gln Ile Thr Cys
 50           55           60
Ser Asp Thr Val Phe Ile Glu Lys Ile Val Leu Ala Phe Ala Ile Leu
 65           70           75           80
Thr Leu Ile Ile Thr Leu Val Cys Val Val Leu Ser Tyr Thr Tyr Ile
 85           90           95
Ile Lys Thr Ile Leu Lys Phe Pro Ser Ala Gln Gln Arg Lys Lys Ala
100           105           110
Phe Ser Thr Cys Ser Ser His Met Ile Val Val Ser Ile Thr Tyr Gly
115           120           125
Ser Cys Ile Phe Ile Tyr Ile Lys Pro Ser Ala Lys Glu Gly Val Ala
130           135           140
Ile Asn Lys Val Val Ser Val Leu Thr Thr Ser Val Ala Pro Leu Leu
145           150           155           160

```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 481 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: rat olfactory epithelium
(B) STRAIN: Sprague-Dawley rat
(F) TISSUE TYPE: olfactory epithelium

(vii) IMMEDIATE SOURCE:

(B) CLONE: J8

(ix) FEATURE:

(A) NAME/KEY: CDS

SUBSTITUTE SHEET

-70-

(B) LOCATION: 2..481

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:21:

C ATC TGC CAC CCG CTC CAC TAC TCT CTT CTC ATG AGT CCT GAC AAC	46
Ile Cys His Pro Leu His Tyr Ser Leu Leu Met Ser Pro Asp Asn	
1 5 10 15	
TGT GCT GCT CTG GTA ACA GTC TCC TGG GTG ACA GGG GTG GGC ACG GGC	94
Cys Ala Ala Leu Val Thr Val Ser Trp Val Thr Gly Val Gly Thr Gly	
20 25 30	
TTC CTG CCT TCC CTC CTG ATT TCT AAG TTG GAC TTC TGT GGG CCC AAC	142
Phe Leu Pro Ser Leu Leu Ile Ser Lys Leu Asp Phe Cys Gly Pro Asn	
35 40 45	
CGC ATC AAC CAT TTC TTC TGT GAC CTC CCT CCA TTA ATC CAG CTG TCC	190
Arg Ile Asn His Phe Phe Cys Asp Leu Pro Pro Leu Ile Gln Leu Ser	
50 55 60	
TGC TCC AGC GTC TTT GTG ACA GAA ATG GCC ATC TTT GTC CTG TCC ATC	238
Cys Ser Ser Val Phe Val Thr Glu Met Ala Ile Phe Val Leu Ser Ile	
65 70 75	
GCT GTG CTC TGC ATC TGT TTC CTC CTA ACC CNN NNN TCC TAC ATT TTC	286
Ala Val Leu Cys Ile Cys Phe Leu Leu Thr Xaa Xaa Ser Tyr Ile Phe	
80 85 90 95	
ATA GTG TCC TCC ATT CTG AGA ATC CCT TCC ACT ACC GGC AGG ATG AAG	334
Ile Val Ser Ser Ile Leu Arg Ile Pro Ser Thr Thr Gly Arg Met Lys	
100 105 110	
ACA TTT TCT ACA TGT GGC TCC CAC CTG GCC GTG GTC ACC ATC TAC TAT	382
Thr Phe Ser Thr Cys Gly Ser His Leu Ala Val Val Thr Ile Tyr Tyr	
115 120 125	
GGG ACC ATG ATC TCC ATG TAT GTC GGC CCA AAT GCG CAT CTG TCC CCG	430
Gly Thr Met Ile Ser Met Tyr Val Gly Pro Asn Ala His Leu Ser Pro	
130 135 140	
GAG CTC AAC AAG GTC ATT TCT GTC TTC TAC ACT GTG ATC ACC CCA CTA	478
Glu Leu Asn Lys Val Ile Ser Val Phe Tyr Thr Val Ile Thr Pro Leu	
145 150 155	
CTG	481
Leu	
160	

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 160 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Ile Cys His Pro Leu His Tyr Ser Leu Leu Met Ser Pro Asp Asn Cys

SUBSTITUTE SHEET